

# Volumetric Hadamard Normalization for Sub-Bottom SAS ATR

Gregory D. Vetaw<sup>1</sup>, Benjamin Cowen<sup>2</sup>, Daniel C. Brown<sup>2,3</sup>,  
David P. Williams<sup>2</sup>, Suren Jayasuriya<sup>1</sup>

<sup>1</sup>School of Electrical, Computer and Energy Engineering; Arizona State University

<sup>2</sup>The Applied Research Laboratory, <sup>3</sup>Graduate Program in Acoustics; Pennsylvania State University

**Abstract**—Image normalization strategies for three-dimensional (3-d) synthetic aperture sonar (SAS) is a relatively underexplored area for automatic target recognition (ATR) leveraging convolutional neural networks (CNNs). For 3-d SAS data, underwater sub-bottom targets (i.e. objects that are buried, or partially buried, in submerged sediment) typically exhibit weaker acoustic returns than the frequently-cluttered sediment bottom, often disrupting conventional dynamic range (DR) compression algorithms (such as log-compression) that use the brightest returns to normalize the rest of the data. This can reduce contrast of the target relative to clutter features, damaging ATR performance. This work proposes a multiplicative preprocessing layer via the element-wise Hadamard product, attached to the beginning of a CNN architecture, that is trained to emphasize salient target features by enhancing the contrast of the input before 3-d features are extracted via convolutional layers. The preprocessing layer is trained with a novel template-matching based regularization function, learning to improve ATR performance. Further, by visualizing the output of this trainable layer, we can interpret what the network considers to be salient features of targets. We train and validate our method on measured volumetric SAS data captured by the Sediment Volume Search Sonar (SVSS) system, achieving an 8% increase in the AUC-PR score when training on a heavily imbalanced dataset. We conduct several ablation studies to validate the design choices of our method. This work shows that learning a transformation that normalizes (or enhances the contrast) sub-bottom SAS data cubes on a pixel-level can produce sub-bottom normalized data that achieves better ATR performance.

**Index Terms**—Synthetic Aperture Sonar, Template-Matching, ATR, Buried Objects, CNN, Dynamic Range Compression

## I. INTRODUCTION

Deep ATR algorithms designed for underwater acoustic vision research have had many successes with improving the classification performance for 2-d SAS imagery using CNNs [1]–[5] as opposed to implementing the traditional feature-based ATR classifiers [6]–[12]. This line of work has inspired SAS ATR developers to employ higher-dimensional CNNs to detect and classify objects buried in the sub-bottom [13], [14] with the recent release of 3-d SAS datasets [15], [16] designed for buried object detection. However, despite CNNs being the preferred SOTA method for several different binary classification tasks on real measured 2-d SAS datasets with objects above the seafloor, incorporating the technology to classify 3-d volumetric SAS imagery of

buried objects still remains an open challenge in terms of lowering the false alarm and false negative rates.

One possible reason why 3-d CNNs tend to struggle with objects buried in the sediment as opposed to surficial objects in water is because insonifying the sediment can produce strong reverberated sound levels that tend to be higher than the reflected echoes from the buried object of interest [17]. Thus, SAS imagery in the sub-bottom have shown low signal-to-noise ratio (SNR) on buried objects. This effect becomes more prominent in very shallow water environments where the underwater imaging sensor tends to capture strong reverberated signals from either multipath scattering or sources originating from the seafloor [18]–[21], specifically from embedded rocks or elevated sound attenuation in the sediment [15] with both sources being labeled as clutter.

Generally, training a 2-d CNN-based SAS ATR algorithm requires that the imagery be normalized to some fixed range (e.g. [0,1] or [-1,1]) prior to being placed as input to the model. However, applying a naive normalization scheme that scales data relative to a global or local maximum value can cause the network to mistakenly associate bright clutter features with labeled targets during training. Further, due to the expense of collecting underwater SAS data, datasets typically have limited target samples available to statistically remove these misleading associations that contribute to a missed detection/classification of a whole target. Thus, the particular method chosen to normalize or transform the volumetric sub-bottom SAS data for 3-d CNN processing plays an important role on how the model makes its predictions. Regularizing the normalization to respond to local changes on DR-compressed data has been shown to reduce the artifacts on a voxel-level and produced volumetric SAS imagery that is suitable for humans to detect the targets in the sub-bottom [15].

In this paper we introduce a data-driven preprocessing layer for 3-d SAS CNN classification of sub-bottom data. Our method learns a transformation that normalizes the entire data cube on a voxel-level by performing Hadamard element-wise products with the incoming data before CNN processing. Allowing the normalization to be learned simultaneously with the CNN to classify buried objects can yield enhanced performance. To the best of our knowledge this area of ATR research is under-explored in the literature and we intend to address this gap.

We also introduce a novel template-matching technique from computer vision to regularize the learned weights of

the normalization layer and optimize it concurrently with the classification error. Given scenes centered on the targets, our method can learn a preprocessing function that can improve CNN-based ATR performance and provides the user a way to interpret the network's decision on what it considers to be salient features of the targets.

Our specific contributions include: (1) a new machine learning layer to increase CNN-based image classification performance on volumetric sub-bottom SAS data with buried and proud objects; and (2) a novel template-matching based regularization function that measures how well the layer is capturing high salient features related to the objects.

To validate the method, we compare against a set of global baseline methods that make the volumetric SAS imagery around the object visually appealing by enhancing contrast of target features. Our results highlight the effectiveness of the learned Hadamard normalization for 3-d SAS ATR. This work is an extension of an earlier conference paper [22]. We have extended the paper's results by implementing new methods to regularize the weights of the normalization layer and mitigate voxel outliers. Additionally, we systematically evaluate our proposed approach with three different CNN architectures.

Our paper is organized as follow. Section II introduces the technical background related to our work and how normalization have been explored in 2-d and 3-ds SAS ATR. Section III introduces the overall pipeline that shows the end-to-end ATR pipeline with proposed method. Here we introduce our volumetric multiplicative Hadamard normalization layer and how we regularize the learned weights using a template-matching loss function. Section IV discusses details pertaining to the SVSS dataset and the training implementation. Section V covers the experimental results quantitatively and qualitatively and also show the ablation studies performed in the design decisions in our machine learning approach. Finally we summarize our work in Section VI, provide the limitations of our method and future works. This work is an extension to a preliminary version which was presented at the International Geoscience and Remote Sensing Symposium (IGARSS) [22].

## II. RELATED WORK

**Deep CNNs for 2-d SAS ATR:** Deep Convolutional Neural Networks (CNNs) typically leverage supervised learning to train on labeled 2-d SAS datasets to achieve superior ATR performance for target detection and object recognition tasks. Some 2-d SAS ATR applications include mine countermeasures (MCM) and the safe clearance of unexploded ordinance (UXO) in the underwater environment. Williams [1] used deep CNNs to discriminate and classify target and non-target imagery using 2-d SAS datasets collected by the CMRE's MUSCLE AUV. Tiny CNNs were introduced and shown that they can reach or even surpass human expert performance with smaller amounts of training data [2]. Galusha et al. found CNNs plus a prediction fusion method deemed viable in classifying SAS image tiles [5]. Chen and Summers [4] used CNNs for semantic segmentation tasks such as separating target from background across several different ocean environments. Gerg and Monga [3] enforced structural priors as regularization

terms to enhance their CNN learning and achieved state-of-the-art performance.

**Deep CNNs for 3-d SAS ATR:** While there has been much recent success with CNNs for 2-d SAS ATR, 3-d SAS ATR is relatively underexplored especially for buried object detection and classification. Williams and Brown built a 3-d CNN detector to find the position of objects in localized regions above and below the sediment [13] by extending the 2-d Mondrian detector [23]. However, performance was limited on buried targets due to embedded rocks and elevated sound attenuation in the sediment which lowered the SNR. Further, the work was extended to include a downstream 3-d CNN-based classifier to predict the labels of centered and isometric data cubes of buried and partially buried targets [14]. Hoang et al. coupled an ensemble of 2-d deep neural networks for improved classification [24], [25]. However, classifying volumetric SAS imagery in shallow-water environments with strong levels of reverberation from background clutter still remains a difficult task for ATR networks, especially when the SAS dataset is severely imbalanced and the SNR on buried objects is low.

**Normalization and Dynamic Range Compression in Machine Learning:** Input normalization is typically applied at the preprocessing stage prior to CNN training as either a basic linear transformation, a histogram-based method, or a whitening-based method which are all summarized by Sendjasi et al. [26]. However we note that there is no standardized normalization procedure that has been established in general for machine learning algorithms including CNNs.

The DR of an image is defined as the ratio of the intensity value of the brightest pixel to the intensity value of the darkest pixel. Methods to reduce the brightness or contrast of an image include linear tone-mapping operators (TMOs) that map an image with a high dynamic range or color scale to a lower dynamic range [27]. There are also nonlinear TMOs that compresses the high end of the intensity distribution when applied on high dynamic range images. One common dynamic range compression (DRC) approach found in the sonar community is the logarithmic-compression (log-compression) function specifically the logarithm with base-10. A study conducted by Buckler et al. showed the importance of dynamic range compression for subsequent CNN performance for a variety of common computer vision tasks [28].

**Normalization for 2-d SAS ATR:** The amplitude of raw complex sonar data with real and imaginary pixels usually exhibit a high DR due to *bright* acoustic scattering off objects, environmental clutter, and the sediment. Compressing the DR, as mentioned earlier, is an effective method to visualize 2-d SAS images with targets on the seafloor. The low dynamic range or DR-compressed SAS imagery generally displays a high resolution image highlighting the target, shadow, and the ocean environment.

Normally, 2-d SAS ATR algorithms leverage single-look complex imagery of targets after image reconstruction for ATR tasks with nearly all methods discarding the phase information and using the magnitude representation alone for processing [29]. It is this representation of the data that gets DR-compressed to the color range of about 40 dB for human

viewing [30] and then normalized to the range of [0, 1] after applying a simple pixel-scaling method.

Usually if the target's signatures are the brightest pixels along with its shadow's features being the darkest after applying the DRC, then the simple pixel-scaling normalization will contribute to the ATR algorithm obtaining good results. However, if there are spatial inhomogeneities or bright point responses found in the background pixels due to bright point scatterers reflecting off the seafloor texture or strong reverberated signals from background clutter then applying a global pixel-rescaling method such as min-max normalization or standardization could lead to the CNN learning to detect the outliers instead of the weaker signal of the target.

To account for spatial changes in the reverberation power level, sonar researchers have designed adaptive normalizers which are modified versions of the global feature scaling method. This approach re-scales the original image by an estimate of the local scattering strength of the background. This derived quantity is used to equalize out the effects caused by bright scatterers [15]. However applying this approach without regularizing the background estimator can introduce artifacts in regions where the estimate is small. To capture large scale artifacts from background clutter, Dobeck [31] used a Fourier-based denoising algorithm. Carbone et al. [32] developed a background normalizer based on the minimum variance spectral estimator (MVSE) which is a power spectral density (PSD) estimator that can be applied in multiple dimensions. Williams [23] designed a two-pass median normalization method before the data is DR-compressed. Note that all of these methods were fixed and could not be learned or optimized end-to-end with the downstream CNN-based ATR algorithm.

**Normalization in 3-d SAS ATR:** As for 3-d SAS, DR-compressing the image of objects buried in the sub-bottom often leaves the target obscured in the imagery by sediment rocks, methane gas bubbles, and other visual clutter making it nearly impossible for humans to detect and label targets when visualized. One avenue to correct the SNR of the 3-d SAS imagery is to enhance the contrast by mapping the clutter pixels to the background or lowering the intensity values of non-target signatures in the imagery. Brown et al. [15] used a regularized normalization based on an estimate of the spatially-varying seabed scattering strength. The method leveraged the split window normalizer, found in detection signal processing, and extended it to 3-d. The method also applied an additional quantile normalization method to remove voxel outliers and then applied both logistic and gamma tone-mappings. They found that the method increased the intensity of the sub-bottom returns that includes buried objects and late-time returns of the objects when maximum intensity projections (MIPs) were taken [33]. Williams and Brown [13] designed a custom volumetric normalization strategy for sub-bottom SAS data by including the 2-d background median normalizer [13] with a stage for determining and removing regions contaminated by high levels of noise reverberation. The normalization approach was applied as a preprocessing algorithm before feeding the data to a CNN-based detection algorithm. In contrast to these previous methods, we let our

CNN-based ATR algorithm learn a volumetric normalization strategy that aids in the classification performance on buried objects.

### III. METHOD

In this section, we present our proposed volumetric Hadamard normalization (VHN) layer for CNN-based ATR classification on sub-bottom SAS data and an analysis of its performance. We also discuss how we regularize the weights of the learned layer and how to mitigate processing artifacts from appearing in the normalized 3-d SAS imagery. Note that our method is designed to be complementary to the choice of CNN-based ATR network deployed, and thus our goal is to show relative performance improvement with the VHN layer.

#### A. Overall Pipeline

Fig. 1 shows the proposed end-to-end ATR pipeline for CNN-based classification on sub-bottom SAS data. The pipeline starts off by computing the amplitude of raw complex sub-bottom SVSS volumetric imagery (after beamforming) and then compresses the DR of the data cube. We utilized log-DRC to perform this compression as it is the most commonly used in the sonar community and highlights features of the target to be learned. A 2-d slice is extracted from each representation in the processing chain to highlight how each step transforms the SAS imagery of a partially buried object. Note the amplitude data is a high dynamic range image with little clutter, yet the target characteristics are discernible. The dynamic range compressed version shows the target surrounded by sediment clutter. The normalized version removes the clutter and preserves the salient features of the target for CNN processing.

Subsequently, we selectively enhance the contrast of the DR-compressed version of the data prior to feeding it as an input to the CNN-based ATR algorithm. This is important for reducing several artifacts that appear in the imagery after placing the sediment clutter and target features on similar color levels from the strict DRC. In this paper, we propose a number of contrast enhancement methods that can increase performance (see Section IV-C for a list of methods we compare). However for optimal performance, we introduce a learnable volumetric normalization layer based on a Hadamard product that can be trained end-to-end based on the backpropagation of CNN performance and a novel template-matching regularization function defined in Section III-C.

#### B. Learned Volumetric Normalization Layer

In this work, we designed a multiplicative volumetric normalization layer to process DR-compressed sub-bottom data cubes. This approach includes: (1) learning a spatially-varying cube of differentiable weights to be applied across the DR-compressed data cube. We call this step a volumetric Hadamard normalization (VHN) with imagery that resembles outputs of a contrast enhancement operation. (2) Regularizing the differentiable cube by penalizing the distance from the normalized average of target cubes in a training

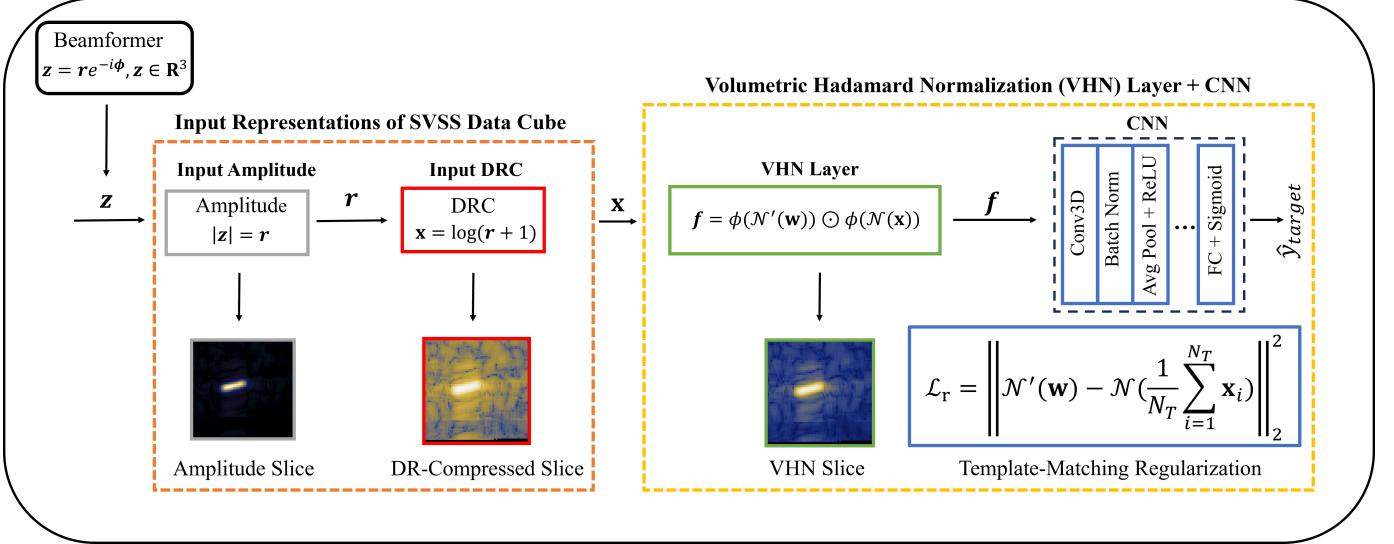


Fig. 1. Proposed pipeline for CNN-based ATR classification. Here full normalized amplitude DR-compressed SVSS data cubes are processed in the volumetric Hadamard normalization (VHN) layer before being classified as a target or clutter. The VHN layer and the CNN are jointly optimized based on the classification error. The learned cube of weights  $\mathbf{w}$  is regularized by a template-matching  $\ell_2$  loss function  $L_r$  where the template is the normalized average of target cubes used in a small batch of training examples. Notice how the DR-compressed imagery in the red box shows target highlights, target resonant scattering, sediment clutter. The VHN layer attempts to suppress the clutter around the target by enhancing the contrast of the imagery in way that preserves the salient features of the target.

batch via the  $\ell_2$  loss function. The combination of DRC and Hadamard normalization results in a composite function, e.g.  $(f \circ g)(x) = f(g(x))$ , which reduces the brightness of clutter pixels around the target (see Fig. 6 exemplary cube 1 MIPs). The advantages of this preprocessing procedure are two-fold: (1) it transforms the data from raw, which usually has a very high dynamic range, to a more suitable version for downstream ATR processing and (2) it removes visual clutter in a way that makes the tone-mapped imagery more appealing to the human visual system.

The VHN layer is defined as follows:

$$f_{VHN}(\mathcal{N}(\mathbf{x}); \mathcal{N}'(\mathbf{w})) = \phi(\mathcal{N}'(\mathbf{w})) \odot \phi(\mathcal{N}(\mathbf{x})). \quad (1)$$

where  $\mathbf{x}$  is DR-compressed data,  $\mathbf{w}$  is learnable weights of the VHN layer,  $\odot$  refers to Hadamard element-wise multiplication [34],  $\mathcal{N}(\mathbf{x}) = \frac{\mathbf{x} - \min(\mathbf{x})}{\max(\mathbf{x}) - \min(\mathbf{x})}$  is min-max normalization, and  $\phi(\cdot)$  is the non-linear, rectified linear unit (ReLU) activation function.  $\mathcal{N}'(\mathbf{w})$  is an adjusted min-max normalization [35] for the weights given by the following equation:

$$\mathcal{N}'(\mathbf{w}) = \frac{\mathbf{w} - \min(\mathbf{w}) + 1}{\max(\mathbf{w}) - \min(\mathbf{w}) + 2}, \quad (2)$$

Eq. 1 takes in min-max normalized DR-compressed SVSS data cubes  $\mathcal{N}(\mathbf{x})$  to perform Hadamard element-wise multiplication  $\odot$  with an adjusted min-max normalized 3-d cube of weights  $\mathcal{N}'(\mathbf{w})$  equal to the size of the input cube. The Hadamard product allows the normalization layer to generalize well when applied on the target cubes drawn from the test set.

Eq. 2 is an adjusted normalization derived by adding 1 and  $-1$  to the maximum and minimum values respectively (similar to [35]). This version of the min-max normalization was implemented over the standard min-max feature normalization to bias the learned  $\mathbf{w}$  to the range  $[0.05, 0.95]$  once training

was complete and to avoid division by zero when the cube stores weights of equal value i.e  $\max(\mathbf{w}) = \min(\mathbf{w})$ . We found that using this approach to normalize the cube of weights helped with the perceptual quality of the tone-mapped imagery and boost model performance which is shown in V-D.

The ReLU activation function  $\phi(\cdot) = \max(0, \cdot)$  [36] was applied on  $\mathcal{N}(\mathbf{x})$  and  $\mathcal{N}'(\mathbf{w})$  before applying the Hadamard product so that the errors in the backward propagation are dependent on  $\mathbf{w}$  and not on the input  $\mathbf{x}$  when applying the chain rule in the backward pass [37]. Combining ReLU with Eq. 2 to normalize  $\mathbf{w}$  allows for non-zero gradient updates to flow to the learned weights. For example, when elements in  $\mathbf{w}$  are zero, Eq. 2 applies a bias in the normalization to avoid the numerical influence of  $\text{ReLU}(0)$  [38] in the backpropagation. Table IV shows how applying different combinations of ReLU and normalization on  $\mathbf{w}$  and  $\mathbf{x}$  effects the model performance. In summary, applying all of the operations mentioned in the design of the VHN layer yielded the highest performance.

At the beginning of training, the 3-d weights were initialized to unity to avoid random outliers and Eq. 2 was used to normalize the weights to 0.5 at the beginning of training. Initializing the cube with a random distribution and normalizing  $\mathbf{w}$  from 0 to 1 using the standard min-max normalization  $\mathcal{N}(\mathbf{w})$  caused a decrease in the AUC-PR score ( $= 0.879$ ). The weights were updated via backpropagation based on the loss of the 3-d ATR classifier and then regularized as explained further in the next section. On average the range of  $\mathbf{w}$  fell in the interval of 0.05 and 0.95 after training was complete.

### C. ATR Loss and Regularization Loss Functions

We utilized the Binary Cross Entropy Loss (BCELoss),

$$\mathcal{L}_{bce} = -\log(\hat{y}) - \log(1 - \hat{y}), \quad (3)$$

as the primary loss function for measuring the predicted labels against the ground truth labels with the target class as ( $y = 1$ ) and the clutter class as ( $y = 0$ ).

Additionally, we implemented a novel perceptual regularization function intended to prevent the VHN layer from overfitting on the clutter class when the dataset imbalance ratio is high and to produce enhanced imagery that generalizes well on the target class. This custom loss function penalizes the distance between the weights in Eq. 1 and the normalized average of the DR-log compressed target cubes in a training batch given by the following equation,

$$\mathcal{L}_r = \|\mathcal{N}'(\mathbf{w}) - \mathcal{N}(\bar{\mathbf{x}})\|_2^2, \quad (4)$$

where  $\bar{\mathbf{x}} = \frac{1}{N_T} \sum_{i=1}^{N_T} \mathbf{x}_j$ ,  $\mathbf{x}_j$ 's are DR-compressed target cubes used in a training batch,  $N_T$  is the number of targets in a batch, and  $\bar{\mathbf{x}}$  is the average cube of targets in a training batch.  $\bar{\mathbf{x}}$  can take values outside of the interval [0, 1] so we normalized it from 0, to 1 using  $\mathcal{N}(\bar{\mathbf{x}})$ . Intuitively,  $\mathcal{L}_r$  is a form of template-matching loss function that encourages the learned VHN layer to enhance the contrast in sub-volumes around targets for data cubes that share common salient target features and attempts to confine  $\mathcal{N}'(\mathbf{w})$  to [0, 1]. Eq. 1 can also be viewed as an contrast enhancement function that attempts to reduce the brightness of the clutter pixels while enhancing the target pixels. Note we observed a trade-off with optimizing for ATR performance and optimizing for perceptual quality of the SVSS imagery. When Eq. 4 was minimized the VHN layer produced normalized imagery that looked realistic with few noticeable artifacts. However, this version of data did not allow the model to generalize well as opposed to the imagery that had clear visual artifacts from the normalization.

We used a logarithmic influence function<sup>1</sup> [39]–[41] applied on the penalty term in Eq. 4 to obtain the final regularization term,

$$\mathcal{L}_{r_{log}} = \frac{1}{2} \log_{10}(\|\mathcal{N}'(\mathbf{w}) - \mathcal{N}(\mathbf{x})\|_2^2), \quad (5)$$

to mitigate the artifacts when using the min-max normalization. Using Eq. 5 as the regularization term drastically increased the training time and produced the best set of SVSS imagery for 2-d visualization. However, applying Eq. 5 in the training procedure caused the performance of the ATR model to stagnate on the validation set.

#### IV. DATASET AND IMPLEMENTATION

##### A. SVSS Sub-Bottom Volumetric Dataset

We leverage data collected by the SVSS system designed for buried object detection in the sub-bottom [15]. The Foster Joseph Sayers Reservoir in Pennsylvania was one of the test sites used to demonstrate the effectiveness of the sensor's ability to image proud and buried objects in 2019. The sensor was deployed in two seasons of the year, summer and winter. The height of the lake in the summer season is usually higher than in the winter season due to snowpack runoff in the springtime. The height variation in the lake gave

<sup>1</sup>This is a common influence function found in robust statistics and computer vision to downweight pixel outliers.

TABLE I  
SUMMARY OF TRAINING DETAILS

Configuration	Model Inputs
Optimizer	Adam
Learning Rate	0.002
Batch Size	40 (Balanced)
Cube Size	64x64x101 (x,y,z)
Epochs	50
Library	PyTorch 1.1.0
GPU	GeForce GTX 1080 GPU
CPU	Intel(R) Core(TM) i7-6850K
OS	16.04.7 LTS (Xenial Xerus)
Memory	64 GB

opportunities to demonstrate the SVSS sensor in a shallow water environment (depths < 1 m) and a slightly deeper water environment ( depths < 5 m). Both deep and shallow SVSS datasets were used in our experiments and has been used by several other research papers [13], [14], [25] as part of the Strategic Environmental Research and Development Program (SERDP).

The SVSS dataset we use has two main classes labeled target and clutter. The target class is the minority class and consists of a shot put, aluminum cylinders, steel pipes, and concrete cylinders. Depending on the type, size, or burial depth of the object the strength of the return signals can either show strong or weak acoustic reflections that highlight the object's features when displayed in the sonar imagery. The clutter class is the majority class and consist of cinder blocks, rocks, and other clutter found in the test site. The clutter can also produce strong reverberated acoustic returns with the same intensity strength as the target in the imagery. We utilized image chips generated from the 3-d Mondrian detection algorithm which can detect multiple objects in a target scene and output sub-volumes centered around the targets [13]. For this dataset the clutter class size far outweighs the target class size in this dataset with the clutter set being 20x bigger then the target set. The target set holds 1221 data cubes and the clutter set holds 17103 data cubes. We trained on 855 target cubes and varied the imbalance ratio throughout our experiments to see how imbalancing the dataset effected the normalization layer's generalization performance on the test targets. We test on 244 target cubes and 244 clutter cubes.

##### B. Architecture

We utilized a 3-d SOTA CNN network from Williams and Brown [14] designed for performing binary classification. We modified the CNN model D architecture to include batch normalization layers between the convolutional layers and the pooling layers with an additional convolutional layer before the dense layer, which helped improve performance. We used NVIDIA GeForce GTX 1080 GPU, PyTorch library for deep learning, Adam optimizer with a learning rate of 0.0002 for the global approaches and 0.002 for the ML approaches, and batch sizes of 40 with 20 target cubes and 20 clutter cubes for

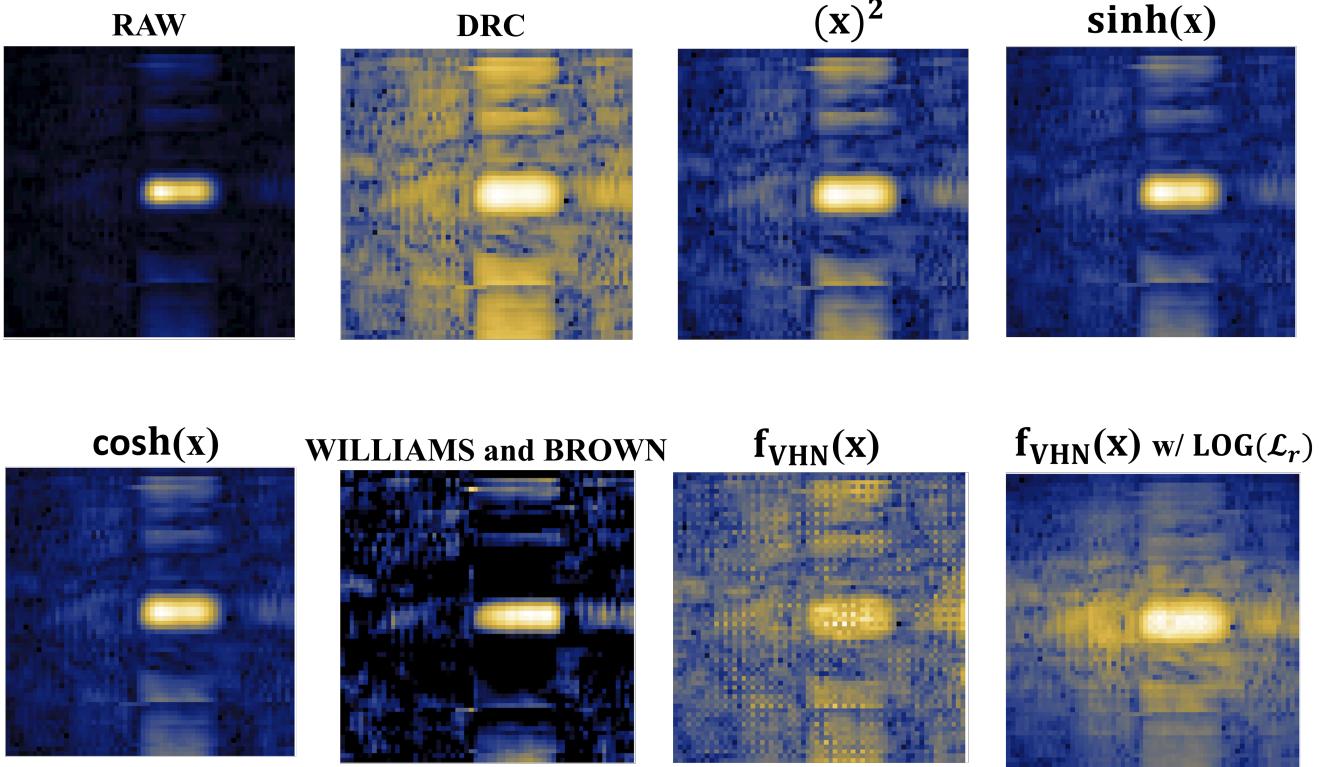


Fig. 2. 8 different representations of a partially buried object captured in the 2019 SVSS dataset. Note how the DRC operation submerges the object in sediment clutter as opposed to the tone-mapped imagery that have an enhanced dynamic range.

training. We updated the CNN and volumetric normalization layer simultaneously and then regularized the weights of the tone mapping function in Eq. 4. A random shuffle of the target list was performed to combat over-fitting on the target class [42], [43]. Each network was trained for 20 epochs with early stopping at 18 epochs. The average time to train was approximately 10 hours for the learnable VHN layer, and 8 hrs with the global functions.

### C. Global Baselines

To study the classification performance and compare how well our volumetric normalization method removes visual clutter from the sub-bottom SAS imagery, we chose several different global functions used as baselines methods to enhance the contrast of the DR-compressed data cubes. Specifically, these functions create a gap between target and clutter voxels by enhancing the intensity of target voxels. We used a normalized DRC version  $\mathcal{N}(x)$ , a squared power law  $(\cdot)^2$ , a hyperbolic sine function  $\sinh(\cdot)$ , a hyperbolic cosine function  $\cosh(\cdot)$ , an exponential function  $\exp(\cdot)$ , and one custom volumetric normalization used for SVSS data [13]. After applying these global contrast enhancement functions, we do min-max normalization on the data as specified earlier in the paper prior to train the CNN models.

Fig. 2 shows the results of normalizing a DR-compressed image of a cylinder through the global functions as well as the VHN layer with and without the logarithmic influence function. The raw amplitude and DR-compressed imagery are also

shown for comparison. Notice that applying a DRC on the raw imagery displays the object of interest surrounded in sediment clutter as opposed to enhancing the contrast afterward DRC. These global contrast enhance functions reduce the clutter and preserves the resonance of the target before CNN processing as shown in the figure. Also, notice how each individual global function maps to a different more appealing visual output than the raw and DR-compressed version.

### D. Performance Metric

To evaluate the classification performance of the network we utilized the area of the curve (AUC) of the Precision-Recall (PR) curve which is a typical performance metric used in machine learning for binary decision problems. We chose the AUC-PR over the the AUC of Receiver Operator Characteristic (ROC) due to the PR being more robust and providing more insights into an algorithm's performance when the dataset has a high imbalance ratio between class distributions [44]. Another advantage of using PR curves is that the model performance can be evaluated for different thresholds as oppose to using a single threshold which is the case with using the accuracy metric to test the performance. We leveraged Scikit-learn metrics package [45] to compute the AUC-PR values.

## V. EXPERIMENTAL RESULTS AND ANALYSES

In this section we cover the main results from implementing the volumetric Hadamard normalization layer with a SOTA

CNN model and compare it against different global baseline approaches for enhancement. We also apply the VHN layer to 3 different CNN-architectures and show how our method helped boost model performance. The clutter-to-target imbalance ratio of the SVSS training dataset is also varied to see how it effects the perceptual quality of the tone-mapped imagery out of the normalization layer and the model performance. We also report the results of different ablation studies used to change various features of the structure of the normalization layer such as with and without the ReLU activation function, with and without the normalization layer, with and without the logarithmic influence function Eq. 5, using the standard and adjusted min-max normalization, and using different activation functions in the design.

**Quantitative Results:** Allowing the CNN network to make predictions based on tone-mapped data as opposed to the traditional raw representation has a direct effect on the classification performance. Table II shows area under the curve for precision-recall (AUC-PR) values from testing on 244 target cubes and 244 clutter cubes after applying 8 different global functions used to enhance the contrast of DR-compressed data cubes. The table shows that using a ML approach yields the highest performance for normalizing proud and buried objects.

TABLE II  
AUC VALUES ON THE SVSS 2019 DATASET  
(2X IMBALANCE)

Data Representation	Precision-Recall Curve
Raw Data	0.837
DRC	0.879
$\mathcal{N}(\mathbf{x})$	0.832
$\mathcal{N}(\mathbf{x}^2)$	0.881
$\mathcal{N}(\sinh(\mathbf{x}))$	0.834
$\mathcal{N}(\cosh(\mathbf{x}))$	0.819
$\mathcal{N}(\exp(\mathbf{x}))$	0.845
Williams and Brown [13]	0.872
$f_{VHN}(\mathcal{N}(\mathbf{x}); \mathcal{N}'(\mathbf{w}))$	<b>0.898</b>

**Qualitative Results:** Figure 2 shows 2-d slices of a partially buried object. The DR-compressed imagery represents the object masked in visual clutter. The global contrast enhance functions attempt to preserve the acoustic energy of the object from the DR-compressed imagery and present a more appealing imagery to human observers than the other forms of the imagery. The VHN layer reduces the clutter on a level of a lower order gamma function but does not preserve the original pixel intensity of the object. This is due to the spatial normalization being applied through the Hadamard element-wise multiplication. Applying a logarithmic influence function on the penalty term as shown in Eq. 5 suppresses the artifacts in the imagery and preserves the brightness of the target and the acoustic resonances. This representation of the imagery would be ideal for observing the features of the object and in highly cluttered environments from the perspective of the human visual system. However adding it to the BCE loss causes the ATR performance to stagnate.

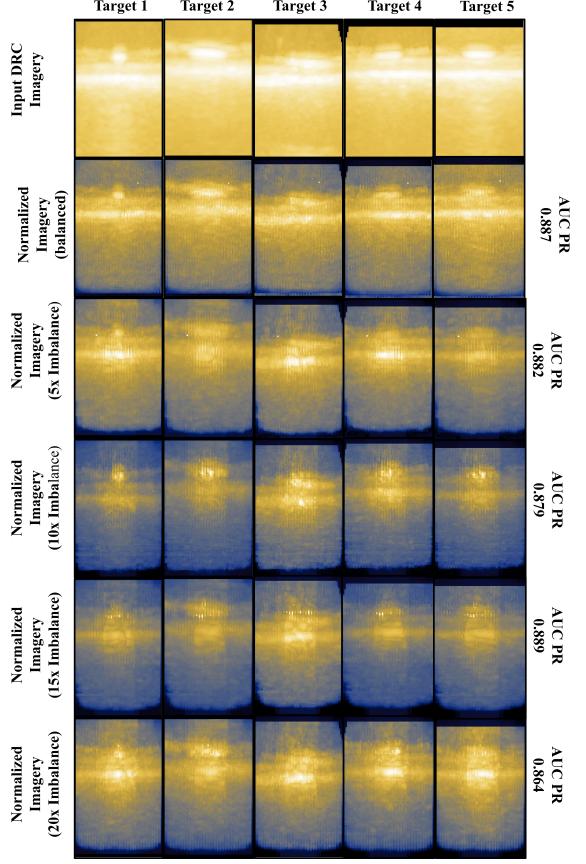


Fig. 3. Shows the along-track MIP of 5 different DR-compressed target examples when processed through the VHN layer. Each row shows the variation of training the layer with different clutter-to-target imbalance ratios. Higher imbalance ratio shows higher extraction of the salient features of the target examples and less clutter in the imagery. 15x imbalance scored the highest AUC-PR value of 0.889. Perceptually 20x produced the best set of imagery but preserved features of the sediment which could be mistaken as a target in the imagery visually and possibly aided to lowering the AUC-PR value which scored 0.864.

#### A. Varying the Clutter-to-Target Imbalance Ratio

Fig. 3 shows the along-track MIP of 5 different DR-compressed target examples when processed through the VHN layer. Each row shows the variation of training the layer with different clutter-to-target imbalance ratios. Higher imbalance ratio shows higher extraction of the salient features of the target examples and less clutter in the imagery. 15x imbalance scored the highest AUC-PR value of 0.889. Perceptually 20x produced the best set of imagery but preserved features of the sediment which could be mistaken as a target in the imagery visually and possibly aided to lowering the AUC-PR value which scored 0.864.

Further a noticeable bright pixel appears in each MIP slice due to applying the min-max linear pixel scaling. As we inject more clutter into the training dataset, the normalization layer is forced to learn more target-class-specific features. At imbalance ratios such as 10x, 15x, and 20x the layer has learned the features that correspond to sediment clutter and the geometry of the sonar beam but still struggles in distinguishing what features belong to the target and what features belong to the interface. As a compromise the layer

learns to slightly reduce the pixels of the target and the pixels that would algorithmically "paint" the sediment into a shape of a target. Future work would be to design a layer that can perform multi-feature normalization or regression so that the target can be tone-mapped separately from the water-sediment interface being that they share same pixel level intensities but different overall geometries.

Fig. 5 shows the results from comparing our method as a preprocessing layer against using only the raw and DR-compressed representations of the data with different clutter-to-target imbalance ratios. The figure shows that applying our VHN layer on the SVSS DR-compressed data helps improves the model performance when the imbalance ratio increases. As mentioned earlier the performance of our approach took a slight dip in the AUC-PR score going from 15x to 20x. However even with this decrease in performance applying the technique still helped boost the model performance from 0.818 (raw) and 0.791 (DRC) to 0.864 when using a 20x clutter-to-target imbalance ratio.

### B. Different Models

We applied our method as a preprocessing layer to 3 other CNN architectures that were designed to classify objects in SAS imagery. We added batch normalization to Williams and Brown CNN Model C [14] and Galusha et al. [5] for training stability and to accelerate convergence [47]. Table III shows the classification performs from implementing the different CNN architectures with and without the VHN layer used as the first layer in the CNN architecture for [14], [5], [46]. As shown in the table, applying the volumetric Hadamard normalization layer drastically increase model performance.

Further, in terms of perceptual quality, Fig. 4 shows the tone-mapped imagery from training our VHN layer with three different CNN designs. The VHN layer was added at the beginning of Williams and Brown [14], Galusha et al. [5], and Xu et al. [46] CNN architectures and applied on a DR-compressed data cube with a proud object. The images show that different CNN designs with VHN layer yields different tone mapped imagery. Overall all training the layer with Williams and Brown CNN design produced the most appealing tone-mapped imagery by reducing the sediment clutter and processing artifacts as well as yielded the highest AUC-PR score of 0.889 on sub-bottom data cubes. Adding the VHN layer with Galusha et al. CNN design did better job at displaying an image with sharp target features such as the edges of the object over adding the layer with Xu et al. but failed to enhance the contrast around the with compared against the other two designs.

### C. Ablation on the Volumetric Hadamard Normalization Layer Design

Designing and training a custom machine learning layer to perform VHN on sub-bottom 3-d SAS data in such a way that that enhances the CNN classification performance and output pleasing and photorealistic imagery is a challenging task, especially when the target and water-sediment interface share similar scattered intensities. We found that model performance

TABLE III  
OUR METHOD WITH DIFFERENT CNN ARCHITECTURES  
(2X IMBALANCE)

CNN Design	Precision-Recall AUC
Model C Williams and Brown [14] + Batch norm	0.838
Model C Williams and Brown [14] + Batch norm + <b>Our Method</b>	<b>0.889</b>
Galusha et al. [5] + Batch norm	0.743
Galusha et al. [5] + Batch norm + <b>Our Method</b>	<b>0.834</b>
Xu et al. [46]	0.752
Xu et al. [46] + <b>Our Method</b>	<b>0.838</b>

Table shows that adding our VHN layer at the beginning of all three CNN models increases performance.

depended on: (1) normalizing both the input data cube and the cube of weights, (2) applying the ReLU activation function after min-max normalization, and (3) using a combination of (1) and (2) in chronological order.

We also found that perceptual quality of the processed imagery out of the learned normalization layer depended on the penalty term used to regularize the weights of the layer. Applying an influence function on the penalty term to downweight the pixel outliers [40] such as the logarithmic function found in Eq. 5 with different bases yielded different results such as model performance and overall perceptual quality of the processed imagery.

To illustrate the effects mentioned above, we conducted an ablation study on the design choice of the volumetric Hadamard normalization (VHN) layer and the choice of logarithmic influence function used on the regularization term described in Eq. 4. We show how all of these design decisions impact the (robustness) of the training, the visual quality, and consistency of the generated images.

Table IV summarizes the result of the ablation study. A key insight that caused the model to perform well with the VHN layer as an learnable preprocessing layer was applying the ReLU activation function on the min-max feature normalized data cube  $\mathcal{N}(\mathbf{x})$  and adjusted min-max normalized cube of weights  $\mathcal{N}'(\mathbf{w})$  before performing the Hadamard product. The ReLU activation function allows for non-zero gradient updates to flow in the backward pass on iterations when  $\mathbf{w}$  is sparse and the adjusted normalization Eq. 2 helps to promote non-zero ReLU derivatives [38] in the backpropagation. Table IV shows that applying the ReLU activation function  $\phi(\cdot)$  subsequently after normalizing the data and the weights  $\phi(\mathcal{N}'(\mathbf{w})) \odot \phi(\mathcal{N}(\mathbf{x}))$  provides an increase in the model performance from 0.846 to 0.890. Also nearly just applying ReLU on the unnormalized weights and data caused an increase in performance from 0.661 to 0.812 which indicates that applying the ReLU function in the VHN layer promotes stability and aids in CNN performance.

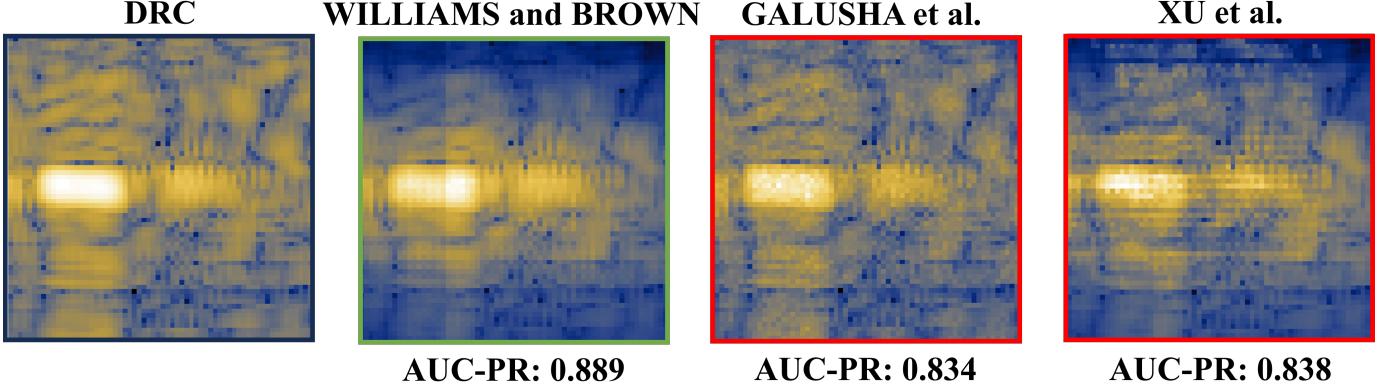


Fig. 4. Shows the perceptual quality from training our VHN layer with three different CNN designs: Williams and Brown [14], Galushash et al. [5], and Xu et al. [46] on a proud object. The images show that different CNN designs with our method yield different tone mapped imagery. Williams and Brown CNN design produced the most appealing tone-mapped imagery by reducing the sediment clutter and processing artifacts as well as yielded the highest AUC-PR score of 0.889 on sub-bottom data cubes.

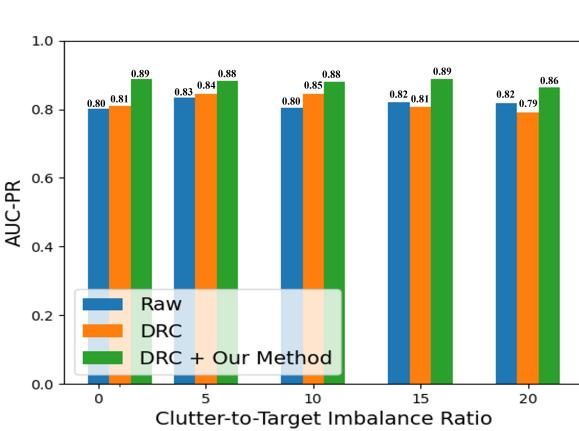


Fig. 5. Shows the AUC-PR scores from varying the clutter-to-target ratio of the SVSS training dataset for different representations of the data as inputs to the CNN. Using an adaptive approach to volumetrically normalize the SVSS data helps improve the model performance when the imbalance ratio is high as opposed to using the raw and DR-compressed representation of the data.

Further, Table IV also shows the results from using different logarithmic influence functions on the MSE regularization loss function. Using a logarithmic influence function to down-weight the voxel outliers provided visually pleasing imagery out of the normalization layer as shown in Fig. 2. However, there is a clear trade-off in optimizing for visual quality versus optimizing for model improvement which is shown in the AUC-PR scores for using the logarithmic influence functions with bases 2 and 10 with a 9% decrease in performance from using approach with just the MSE regularization loss function.

Table IV also shows a study where we explored applying the ReLU function on unnormalized data and weights, applying the ReLU on just the data or just the learned weights, using the template  $\bar{x}$  to perform Hadamard product with  $x$ , and using the ReLU on the output of the VHN layer  $f_{VHN}(\mathcal{N}(x); \mathcal{N}'(w))$ . We found that the highest model performance was obtained with the operations described in Eq. 1 using 2X imbalanced. We chose to perform the remaining experiments described in this

report with this version of the VHN layer design.

TABLE IV  
ABLATION STUDY ON DIFFERENT VOLUMETRIC NORMALIZATION (VHN)  
LAYER DESIGNS  
(2X IMBALANCE)

VHN Design Choice	Precision-Recall AUC
$w \odot x$	0.661
$\mathcal{N}(\bar{x}) \odot \mathcal{N}(x)$ (Template)	0.677
$w \odot \mathcal{N}(x) w \setminus \mathcal{L}_{rl_{log}}$ [Eq. 5]	0.779
$w \odot \mathcal{N}(x) w \setminus \frac{1}{2} \log_2(L_r)$	0.802
$\phi(w) \odot \phi(x)$	0.812
$\mathcal{N}'(w) \odot \mathcal{N}(x)$	0.846
$\mathcal{N}'(w) \odot \phi(\mathcal{N}(x))$	0.847
$\phi(\mathcal{N}'(w)) \odot \mathcal{N}(x)$	0.864
$\phi(f_{VHN}(\mathcal{N}(x); \mathcal{N}'(w)))$	0.888
$\phi(\mathcal{N}'(w)) \odot \phi(\mathcal{N}(x))$	<b>0.890</b>

Table shows that applying the ReLU activation function  $\phi(\cdot) = \max(0, \cdot)$  on  $N(w)$  and  $N(x)$  in the VHN layer design produces the highest AUC-PR value.

#### D. Ablation on the Choice of Normalization

In this work the choice of normalization plays an important role in terms of CNN model performance and in the perceptual quality of the tone-mapped imagery. Table IV shows the AUC-PR score of 0.661 for unnormalized weights  $w$  and input DR-compressed data  $x$  in the VHN design. Normalizing the weights as  $\mathcal{N}'(w)$  and normalizing as  $\mathcal{N}(x)$  produced an increased in performance to 0.846 and applying the ReLU activation function boosted the score to 0.898. One reason for using Eq. 2 in the design choice was to avoid division when the weight cube stores the same value. We investigated using  $\mathcal{N}(w)$  in Eq. 1 with the ReLU activation function and found that the performance decreased slightly to 0.888 from 0.898 from using  $\mathcal{N}'(w)$ . The noticeable difference from using the normalization methods came in the perceptual quality of the tone-mapped imagery out of the VHN layer. Fig. 6 shows

the MIP slice of two DR-compressed target cubes along with a 2-d slice of the actual target. One target slightly buried and another submerged in sediment clutter. The histogram diagrams are shown to highlight how the VHN transforms the intensity distribution of the SVSS images. The highlight of the experiment is found in the first row where the first column (image A) is the MIP slice showing little target features due the loud sediment producing bright voxels. The histogram of the MIP slice shows a skewed distribution to the right of the plot. The second column (image B) is the MIP slice from using the standard min-max normalization which doesn't show any features of the target after VHN tone-mapping. The hisogram shows a noisy pixel distribution. The last column (image C) shows the MIP slice from using the adjusted min-max normalization in the VHN design produced tone-mapped imagery showing features of the target in the maximum intensity plane. This shows that the VHN layer is learning a strategy to tone-map the imagery in a way to reduce the intensity voxels of the clutter and boost the signal related to target features. In the histogram plot the adjusted normalization strategy rescales and shifts the pixel distribution to highlight features of the target in the MIP slice and in the 2-d slice.

#### E. Ablation on the Choice of Non-Linear Activation Function

In our study we discovered that applying the non-linear ReLU activation function on  $\mathcal{N}'(\mathbf{w})$  and  $\mathcal{N}(\mathbf{x})$  in the design of the VHN layer increased the AUC-PR score by well over 3% as opposed to not applying the function in the design. This insight prompt us to study the choice of activation function in the design. Here we chose the hyperbolic tangent function, the leaky ReLU function, the tangent function to measure the relative increases in performance against the ReLU activation function which was the function used in the decision design to perform the experiments in this work. Fig. 7 is a bar chart showing the average AUC-PR score for 5 trial runs using the different activation functions  $\phi(\cdot)$  in Eq. 1 and the baseline design without any activation function  $\mathcal{N}(\mathbf{w}) \odot \mathcal{N}(\mathbf{x})$ . Applying the ReLU activation function in the VHN design achieved the highest average score of 0.889 which supports our decision to use it in the design.

## VI. CONCLUSION

In this paper, we considered the problem of classifying objects buried above and below the sediment in shallow water environments using CNN-based ATR. We developed a differentiable image processing method that can be jointly trained with 3-d CNN models by using a novel target template-matching regularization loss function. We showed that our data-driven image processing method can normalize and tone-map the dynamic range of SVSS data cubes by enhancing the contrast of target features. We also showed that with careful design decisions our method helped with the visual identification of targets in sub-bottom SAS imagery and helped increased the classification performance on buried and partially buried objects using three different 3-d CNN models when performing volumetric Hadamard with incoming data cubes.

We also showed that our method works well when the dataset is highly imbalanced relative to the raw and DR-compressed version of the SVSS data. Thus, training an additional layer to normalize and tone-map volumetric SAS imagery in way that boost model performance can help in the human perception on what is a good image representation to use in the context of using CNN-based models for sub-bottom SAS ATR detection and classification.

However, there were several limitations we encountered with applying the volumetric Hadamard normalization method on a challenging real-world 3-d SAS dataset. First, we noticed that using bad target samples or samples with small objects that have very weak return signals can cause the method to learn a bad normalization strategy where the loud bright sediment is mistaken as a very large target when applying the method on the validation set. Second, we also noticed that the approach could learn a tone-mapping function that over-fits on proud target scenes instead of generalizing across all scenes with buried and proud targets. To overcome this limitation, we had to regularize the cube of 3-d weights in the VHN by using the average DR-compressed target cubes in a training batch as an estimate of what a good template to match and learn target features. Third, our VHN layer is non-adaptive which means that it cannot change its weights with respect to new incoming data cubes after being optimized on the training dataset. Finally, the SVSS data cubes have to be normalized to the range of 0 and 1 for CNN-based processing. While this did optimize the CNN performance, it does yield some visual artifacts when using the VHN layer that seem artificial in nature, a compromise that this paper chose to accept.

While our approach did show success in outputting normalized SVSS imagery that can aid downstream CNN classification on target and clutter data cubes, the approach was contingent on the use of a template-matching regularization loss function which sums all the target cubes in a training batch. Future work could improve this loss function to look at multi-feature normalization that can distinguish between target, clutter, and sediment-water interface. Further, a Hadamard normalization procedure that utilizes supervoxels or contiguous regions of voxels could also help enhance ATR performance by reducing visual artifacts as compared to simple per-voxel normalization. Also adding in more multiplicative VHN layers could support stronger backpropagation when learning to tone-map the target features. We believe that more research is needed to obtain optimal preprocessing and normalization for SAS sub-bottom ATR in general, especially when the end goal is machine vision. We hope that this paper helps inspire the community to rethink the role of dynamic range compression and tone-mapping for such pipelines in the future.

## REFERENCES

- [1] D. Williams, "Underwater target classification in synthetic aperture sonar imagery using deep convolutional neural networks," *23rd International Conference on Pattern Recognition (ICPR)*, pp. 2497–2502, Dec. 2016.
- [2] ———, "On the use of tiny convolutional neural networks for human-expert-level classification performance in sonar imagery," *IEEE J. Oceanic Eng.*, vol. 46, no. 1, pp. 236–260, Jan. 2021.

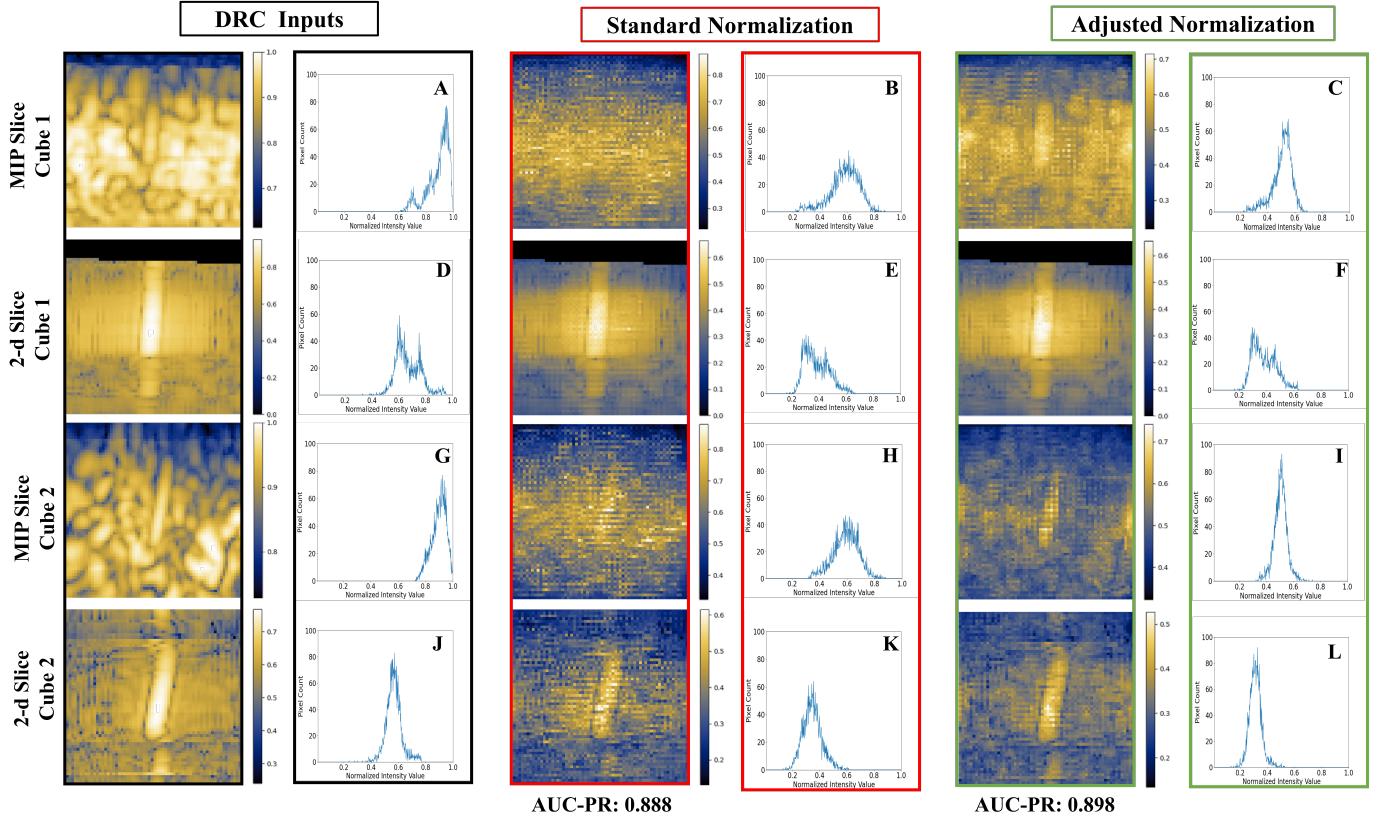


Fig. 6. Shows the MIP slice and the 2-d slice of a target from two separate DRC target cubes along the depth dimension. The first two columns show the DR-compressed input images with the histogram diagrams. The middle two columns show the output imagery from using the standard min-max normalization  $\mathcal{N}(\mathbf{w})$  in the VHN design along with the histogram diagrams. The right two columns show the output imagery from using the adjusted min-max normalization [35]  $\mathcal{N}'(\mathbf{w})$  in the VHN design along with the histogram diagrams. Using the adjusted min-max normalization in the VHN layer produces tone-mapped imagery with features of the target in the maximum intensity plane as shown in image C and B. The histogram diagrams are shown to highlight how the VHN transforms the intensity distribution of the SVSS images. The adjusted normalization strategy rescales and shifts the pixel distribution to highlight features of the target in the MIP slice and in the 2-d slice.

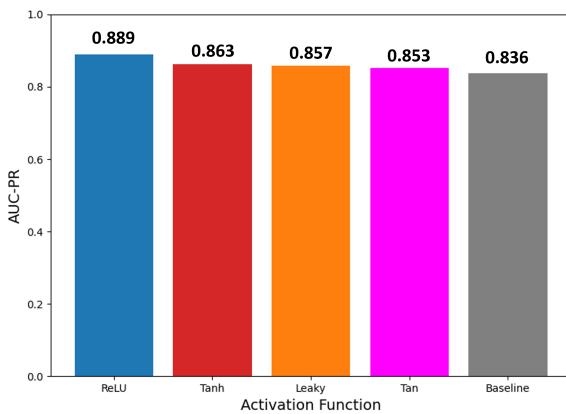


Fig. 7. Shows the average AUC-PR score after 5 trial runs for 4 different activation functions applied separately on  $\mathcal{N}'(\mathbf{w})$  and  $\mathcal{N}(\mathbf{x})$  in the VHN layer design. Using the ReLU activation function in the design achieved the highest average score of 0.889 over the baseline method which uses no activation function.

[3] I. Gerg and V. Monga, "Structural prior driven regularized deep learning for sonar image classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 60, pp. 1–16, Jan. 2022. [Online]. Available:

- <https://doi.org/10.1109/\%2Ffgs.2020.3045649>
- [4] J. Chen and J. Summers, "Deep convolutional neural networks for semi-supervised learning from synthetic aperture sonar (sas) images," *The Journal of the Acoustical Society of America*, vol. 141, pp. 3963–3963, May 2017.
  - [5] A. Galusha, J. Dale, J. Keller, and A. Zare, "Deep convolutional neural network target classification for underwater synthetic aperture sonar imagery," in *Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XXIV, 11012*, ser. Proc. SPIE, vol. 11012, May 2019, p. 5.
  - [6] J. McKay, V. Monga, and R. Raj, "Robust sonar atr through bayesian pose-corrected sparse classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 10, pp. 5563–5576, 2017.
  - [7] N. Kumar, U. Mitra, and S. Narayanan, "Robust object classification in underwater sidescan sonar images by using reliability-aware fusion of shadow features," *IEEE Journal of Oceanic Engineering*, vol. 40, no. 3, pp. 592–606, 2015.
  - [8] R. Fandos and A. M. Zoubir, "Optimal feature set for automatic detection and classification of underwater objects in sas images," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 3, pp. 454–468, 2011.
  - [9] J. C. Isaacs, "Laplace-beltrami eigenfunction metrics and geodesic shape distance features for shape matching in synthetic aperture sonar," in *CVPR 2011 WORKSHOPS*, 2011, pp. 14–20.
  - [10] D. P. Williams and E. Fakiris, "Exploiting environmental information for improved underwater target classification in sonar imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 10, pp. 6284–6297, 2014.
  - [11] T. Fei, D. Kraus, and A. M. Zoubir, "Contributions to automatic target recognition systems for underwater mine classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 1, pp. 505–518, 2015.

- [12] J. Isaacs and J. Tucker, "Signal diffusion features for automatic target recognition in synthetic aperture sonar," in *2011 Digital Signal Processing and Signal Processing Education Meeting (DSP/SPE)*, 2011, pp. 461–465.
- [13] D. Williams and D. Brown, "New target detection algorithms for volumetric synthetic aperture sonar data," *Proc. of Meetings on Acoustics*, vol. 40, p. 070002, Sept. 2020.
- [14] ——, "Three-dimensional convolutional neural networks for target classification with volumetric sonar data," *Proc. of Meetings on Acoustics*, vol. 44, p. 070005, Aug. 2021.
- [15] D. Brown, C. Brownstead, and S. Johnson, "Sediment volume search sonar development: Final report," *DOD Strategic Environmental Research and Development Program (SERDP) Interim Report*, May 2018. [Online]. Available: <https://serdp-estcp.mil/projects/details/97e4a43c-bb65-432d-94ba-6aaff361571a>
- [16] T. Marston, D. Plotnick, and K. Williams, "Three dimensional fast factorized back projection for sub-sediment imaging sonars," in *OCEANS 2019 MTS/IEEE SEATTLE*, Oct. 2019.
- [17] S. Schock, A. Tellier, J. Wulf, J. Sara, and M. Ericksen, "Buried object scanning sonar," *IEEE J. Oceanic Eng.*, vol. 26, pp. 677 – 689, Nov. 2001.
- [18] R. E. Hansen, H. J. Callow, T. O. Sabo, and S. A. V. Synnes, "Challenges in seafloor imaging and mapping with synthetic aperture sonar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 10, pp. 3677–3687, 2011.
- [19] D. A. Cook and D. C. Brown, "Synthetic aperture sonar image contrast prediction," *IEEE Journal of Oceanic Engineering*, vol. 43, no. 2, pp. 523–535, 2018.
- [20] Z. Lowe and D. Brown, "Multipath reverberation modeling for shallow water acoustics," vol. 34, pp. 1285–1291, 01 2012.
- [21] I. Kirsteins, J. Fay, and J. Kelly, "Suppressing reverberation by multipath separation for improved buried object detection," in *MTS/IEEE Oceans 2001. An Ocean Odyssey. Conference Proceedings (IEEE Cat. No.01CH37295)*, vol. 1, 2001, pp. 236–244 vol.1.
- [22] G. D. Vetaw, B. Cowen, D. C. Brown, D. P. Williams, and S. Jayasuriya, "Learning-based tone mapping to improve 3d sas atr," in *IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium*, 2023, pp. 6995–6998.
- [23] D. Williams, "The mondrian detection algorithm for sonar imagery," *IEEE Trans. Geosci. Remote Sensing*, vol. 56, no. 2, pp. 1091–1102, Feb. 2018.
- [24] T. Hoang, K. S. Dalton, I. D. Gerg, T. E. Blanford, D. C. Brown, and V. Monga, "Resonant scattering-inspired deep networks for munition detection in 3d sonar imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–17, 2023.
- [25] T. Hoang, K. Dalton, I. Gerg, T. Blanford, D. Brown, and V. Monga, "Domain enriched deep networks for munition detection in underwater 3d sonar imagery," *IEEE Int. Geosci. Remote Sensing Symposium*, pp. 815–818, July 2022.
- [26] A. Sendjasni, D. Traparic, and M.-C. Larabi, "Investigating normalization methods for cnn-based image quality assessment," in *2022 IEEE International Conference on Image Processing (ICIP)*, 2022, pp. 4113–4117.
- [27] T. Borer and A. Cotton, "A display-independent high dynamic range television system," *SMPTE Motion Imaging Journal*, vol. 125, no. 4, pp. 50–56, 2016.
- [28] M. Buckler, S. Jayasuriya, and A. Sampson, "Reconfiguring the imaging pipeline for computer vision," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 975–984.
- [29] I. Gerg and D. Williams, "Additional representations for improving synthetic aperture sonar classification using convolutional neural networks," *ArXiv*, vol. abs/1808.02868, Aug. 2018.
- [30] J. Groen, E. Coiras, and D. P. Williams, "Detection rate statistics in synthetic aperture sonar images," *NATO STO Centre for Maritime Research and Experimentation (CMRE) Annual Report*, 2009.
- [31] G. J. D., "Adaptive large-scale clutter removal from imagery with application to high-resolution sonar imagery," in *Defense + Commercial Sensing*, 2010. [Online]. Available: <https://api.semanticscholar.org/CorpusID:60674894>
- [32] C. Carbone and S. Kay, "A novel normalization algorithm based on the three-dimensional minimum variance spectral estimator," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 1, pp. 430–448, 2012.
- [33] J. Wallis, T. Miller, C. Lerner, and E. Kleerup, "Three-dimensional display in nuclear medicine," *IEEE Trans. Med. Imag.*, vol. 8, no. 4, pp. 297–303, Dec. 1989.
- [34] G. Stylianopoulos, "Hadamard products and multivariate statistical analysis," *Linear Algebra and its Applications*, vol. 6, pp. 217–240, 1973.
- [35] S. Sinsomboonthong, "Performance comparison of new adjusted min-max with decimal scaling and statistical column normalization methods for artificial neural network classification," *International Journal of Mathematics and Mathematical Sciences*, vol. 2022, pp. 1–9, 04 2022.
- [36] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *International Conference on Machine Learning*, 2010.
- [37] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, pp. 533–536, 1986.
- [38] D. Bertoin, J. Bolte, S. Gerchinovitz, and E. Pauwels, "Numerical influence of  $\text{relu}'(0)$  on backpropagation," *Advances in Neural Information Processing Systems*, vol. 34, 2021.
- [39] P. Huber, *Robust Statistics*. New York, NY: Wiley, 1981.
- [40] R. Brunelli, *Template Matching Techniques in Computer Vision: Theory and Practice*. Robust Statistics. New York, NY: Wiley, 1981.
- [41] C.-C. Chuang, J.-T. Jeng, and C.-C. Hsiao, "Constructing an influence function of robust learning algorithm based on error distributions for the neural networks," in *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No.02CH37290)*, vol. 3, 2002, pp. 2509–2514 vol.3.
- [42] G. Batista, R. Prati, and M. Monard, "A study of the behavior of several methods for balancing machine learning training data," *SIGKDD Explorations*, vol. 6, pp. 20–29, 06 2004.
- [43] H. He and E. Garcia, "Learning from imbalanced data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, pp. 1263–1284, 2009.
- [44] J. Davis and M. Goadrich, "The relationship between precision-recall and roc curves," in *Proceedings of the 23rd International Conference on Machine Learning*, ser. ICML '06. New York, NY, USA: Association for Computing Machinery, 2006, p. 233–240. [Online]. Available: <https://doi.org/10.1145/1143844.1143874>
- [45] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [46] Y. Xu, X. Wang, K. Wang, J. Shi, and W. Sun, "Underwater sonar image classification using generative adversarial network and convolutional neural network," *IET Image Processing*, vol. 14, no. 12, pp. 2819–2825, 2020.
- [47] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 37. Lille, France: PMLR, 07–09 Jul 2015, pp. 448–456.