

Testing of a composite wavelet filter to enhance automated target recognition in SONAR

Jeffrey N. Chiang, USRP Summer Intern
University of California, Los Angeles, Los Angeles, CA 90095

Mentor: Dr. Thomas T. Lu
Jet Propulsion Laboratory, Pasadena, CA 91109

Automated Target Recognition (ATR) systems aim to automate target detection, recognition, and tracking. The current project applies a JPL ATR system to low resolution SONAR and camera videos taken from Unmanned Underwater Vehicles (UUVs). These SONAR images are inherently noisy and difficult to interpret, and pictures taken underwater are unreliable due to murkiness and inconsistent lighting. The ATR system breaks target recognition into three stages: 1) Videos of both SONAR and camera footage are broken into frames and preprocessed to enhance images and detect Regions of Interest (ROIs). 2) Features are extracted from these ROIs in preparation for classification. 3) ROIs are classified as true or false positives using a standard Neural Network based on the extracted features. Several preprocessing, feature extraction, and training methods are tested and discussed in this report.

I. Introduction

Automated target recognition (ATR) has been a focus of image processing and artificial intelligence research for some time, with immediate applications in surveillance and autonomous navigation. There are several approaches to ATR, each with limitations in performance and resources. It is very difficult to develop a generalized ATR algorithm due to the complexity of targets and background in various real-world applications. NASA-JPL has designed a multi-stage ATR framework that can be modified and optimized for different inputs. First, an image is de-noised and regions of interest (ROIs) are detected using various image filtering techniques, then a feature extraction is performed on these ROIs, and finally, a neural network is used to classify each ROI as a target or a false positive. The goal of the internship project was to apply this framework and to optimize it for use with low resolution SONAR and camera videos taken by an unmanned underwater vehicle (UUV). Different filter and feature extraction techniques are explored to generate the highest detection rate possible while keeping a low false positive count.

II. Background

2.1 ATR Framework

The multi-stage ATR framework developed at NASA-JPL generally breaks image processing into three steps. These steps can be summarized as preprocessing, feature extraction, and false positive reduction, as shown in Figure 1. In this way, speed and accuracy can be balanced. Generally, computationally expensive processing techniques such as filtering and neural networks must be optimized to achieve real-time operations in a computer. The ATR framework aims to identify ROIs that could contain targets in the first step. These ROIs are passed into the neural network for classification--which allows for much faster computation time when compared to analyzing the entire image.

In this framework, preprocessing is defined as any transform performed on the image before features are extracted. Typically, the raw image can be normalized using a histogram filter, in order to enhance contrast. To identify targets, a target filter is constructed, either using a simple image editing program or a wavelet function. The image and filter are correlated in the Fourier domain, and a threshold is applied to determine ROIs. This step is responsible for the upper limit of the system's performance--that is, if a target is not detected in this step, it will

never be detected by the ATR system. Therefore, it was important to keep the threshold relatively low, to increase the likelihood that the actual target is present in the ROIs detected.

While preprocessing and ROI detection determine percentage of targets actually "caught" by the system, feature extraction and the classification algorithm are essential for the false positives eliminated. The goal of feature extraction is twofold: first, to provide essential features that will help the classifier differentiate between true targets and false positives, and second, to reduce the dimensionality of the ROIs to cut computation time. In this project, features were selected in an attempt to quantify our own observations and classifications of the images. As will be shown later, the extracted features are mostly responsible for the average number of false positives per image.

Finally, a reliable classification algorithm must be used to classify ROIs as targets or false positives. Typically a gradient descent back-propagation neural network has been used in the ATR system since it is producible in hardware. These neural networks employ supervised learning algorithms. Thus it was necessary to first train the network on sample data, described in detail in the methods section.

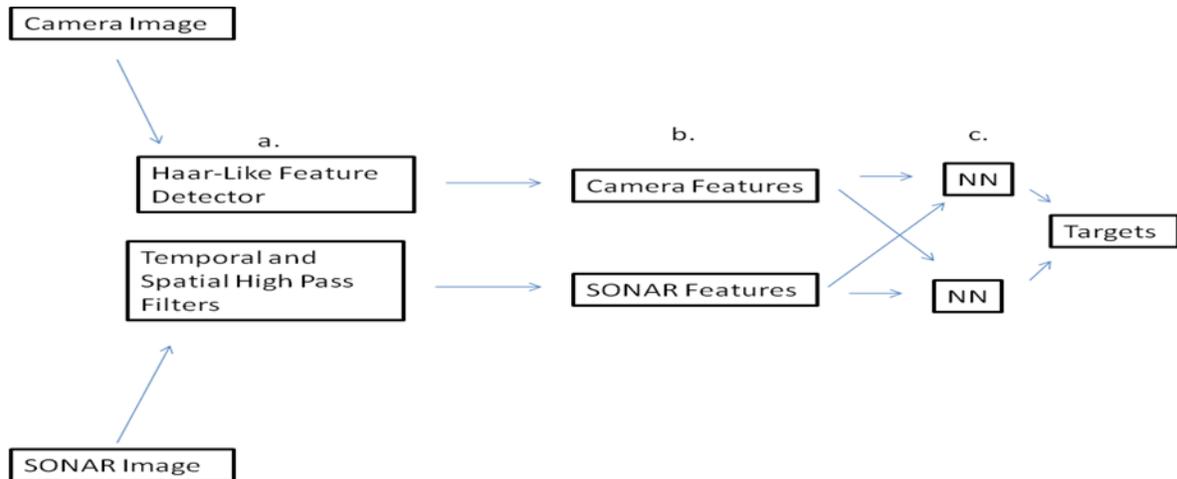


Figure 1. Representation of ATR System. In a) ROI detection occurs using a Haar-Like Feature detector for camera images, and composite filters for SONAR images. b) Features are extracted from the detected ROIs, correlated with each other and passed into neural networks for classification (c).

2.2 Video Dataset

The SONAR images used in this study were extracted from videos taken by UUVs. Because of the nature of sound waves, SONAR images are visualized as arcs, shown in Figure 2. The UUV also has a video camera that takes continuous video images of the underwater target. The camera images are generally very low quality, as shown in Figure 3(a). Haar-like filters were applied to enhance the image quality, as shown in Figure 3(b).

We were provided with three video data sets, each with SONAR images and camera images. Previous work had been done by Zhang et al. [5] using Haar-like features to analyze the camera images. Thus we aimed to generate a robust ATR system for SONAR that was reliable both across ranges (within each video), and across the three videos, which would complement the camera ATR system.

For ease of processing, each SONAR image was converted to a 64 x 640 rectangular image, using a process analogous to converting polar coordinates to Cartesian space:

The SONAR images provided were also taken at five different ranges, in which the targets appear very different from each other.

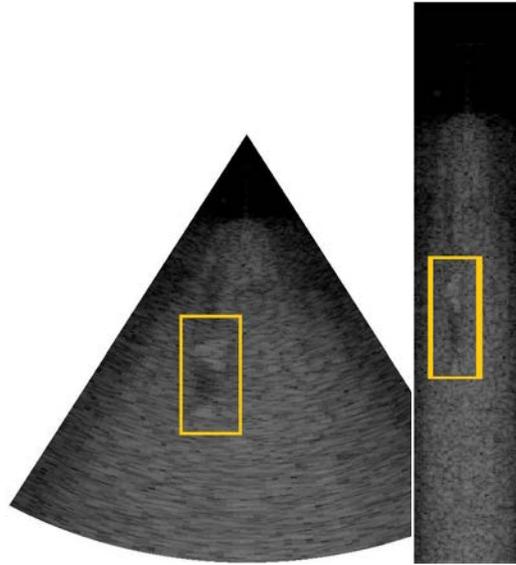


Figure 2. Converted sonar image example. Target is labeled in a yellow bounding box.

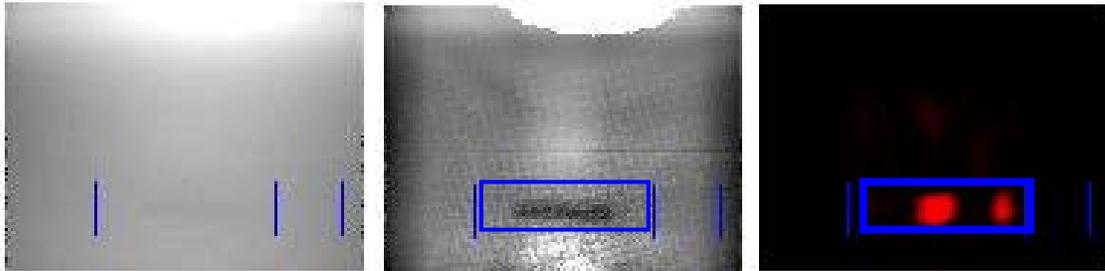


Figure 3. Camera image processing using Haar-like features. This process is described in Zhang and Lu [5]. The output from this analysis is correlated with features extracted from SONAR images to improve false positive reduction.

III. Methods

3.1 Preprocessing using the Mexican-Hat Wavelet Filter

In image processing, the Mexican-Hat wavelet is a commonly used filter for edge and blob detection. The function is sometimes approximated by a difference of two Gaussian curves, but in this study the following was used:

$$\varphi(t) = \frac{2}{\sqrt{(3\sigma)\pi^{\frac{1}{4}}}} \left(1 - \frac{x^2}{\sigma^2}\right) e^{-\frac{x^2}{2\sigma^2}} \quad (1)$$

This wavelet, as shown in Figure 4, is an intuitive choice for detecting targets in the provided SONAR images. Targets and edges corresponding to the wavelet are amplified, while all other pixels are suppressed. However, the size, intensity, and shape of targets vary across different ranges within videos. They also vary substantially across videos, making it unfeasible to use one simple filter in the system.

To remedy the problem of variance within ranges, a composite filter made from two Mexican-Hat wavelets was used, as illustrated in Figure 5. Essentially, the filter searched the image for two differently shaped blobs: a small, bright blob corresponding to the target, and a long dark blob immediately below it corresponding to its shadow. This was implemented as a linear combination of two wavelets, with each given arbitrary weights. Several of these filters were generated, corresponding to a few exemplars arbitrarily picked from the image set. For each range of sonar, five of these composite filters were averaged to create a generalized filter. This totaled to five filters, one for each range of SONAR, eliminating the need to generalize across ranges.

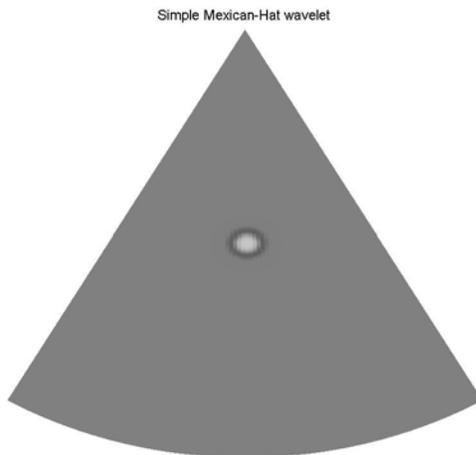


Figure 4. Single Mexican-Hat Blob Filter

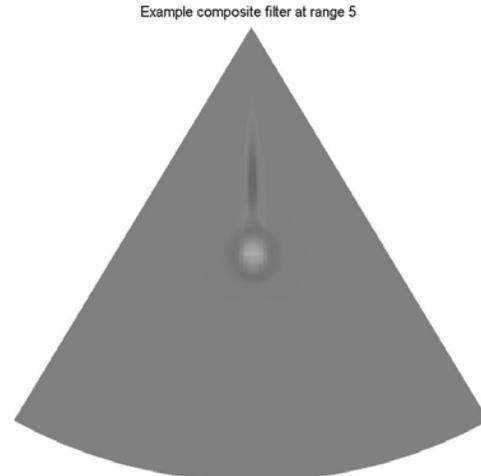


Figure 5. Composite Filter. This is a linear combination of two blob filters, one corresponding to the target as in figure 3, and another corresponding to the shadow.

Images were filtered in the temporal and spatial domains. For the temporal filter, filtered images were compared with the average of the last five images, making the filter sensitive to changes. The spatial domain simply looked for ROIs in the current image. The two filtered images were combined using a weighted average to create the decision image, which determined ROIs.

3.2 Feature Extraction

The seven features chosen to classify ROIs were based on previous ATR work and simple intuition and observation.

1. Distance of the target: The SONAR data provides the range (distance) information. It appeared that targets tended to appear in the middle ranges of the SONAR images.
2. Area of the target: This was found by running a Mexican-Hat filter (designed only for the target) on the raw ROI. The area was represented by the percentage of the ROI covered by the response from the filter. This was chosen because targets are different sizes at various ranges, so an object too big or too small could be dismissed as noise.
3. Ratio of size to distance of target: A close up target would become smaller as it moved further away. Since a neural network deals with linear transformations of its inputs, providing this ratio would emphasize the relationship between area and distance of the target.
4. Peak to side-lobe ratio (PSR): This took the ratio of the highest intensity regions in an ROI to the surrounding area. This was used to represent the intensity and contrast of potential targets.
5. Area of the shadow of the target: A Mexican-Hat filter was used to find the shadow below the ROI, and its area was expressed as a percentage. Since all true targets cast a sound shadow, it was crucial that any ROI with a nonexistent shadow was eliminated as a false positive.

6. *Ratio of width to height of shadow*: The shadows also appeared to have a distinctive size and shape depending on range. This was a rough attempt to quantify the dimensions of the shadow.

7. *Correlation with camera image*: This was a binary variable for whether the azimuth of the target corresponded with the X coordinate in the camera image. If an ROI corresponded with an object detected in the video camera, it was more likely to be a true positive.

3.3 Neural Network Classifier

A three layer feedforward backpropagation neural network is trained to identify the targets. In addition to the seven features outlined above, an additional five binary inputs were included to specify what range the image was taken from. With these dummy variables, the neural network essentially can use a different set of parameters for each range--the equivalent of having a neural network for each of the five ranges. With 12 input neurons, an arbitrarily chosen 11 hidden units was used in the false positive eliminator. The specific neural network used a Levenberg-Marquardt minimization learning algorithm, with random starting weights. Because the starting weights affect the performance of the network, 50 networks were trained and the best one was kept. Performance was defined by entropy [5].

There were several approaches as to how to best train the neural network. For the first set of tests, a neural network was trained for each video, which we will define as an "optimized" neural network for that video. These neural networks were each tested against all three videos, for a total of nine tests. Then, a mixed training set of about 5000 images was hand-picked from the three videos. A network was trained for this set, and it was also tested against the three original videos.

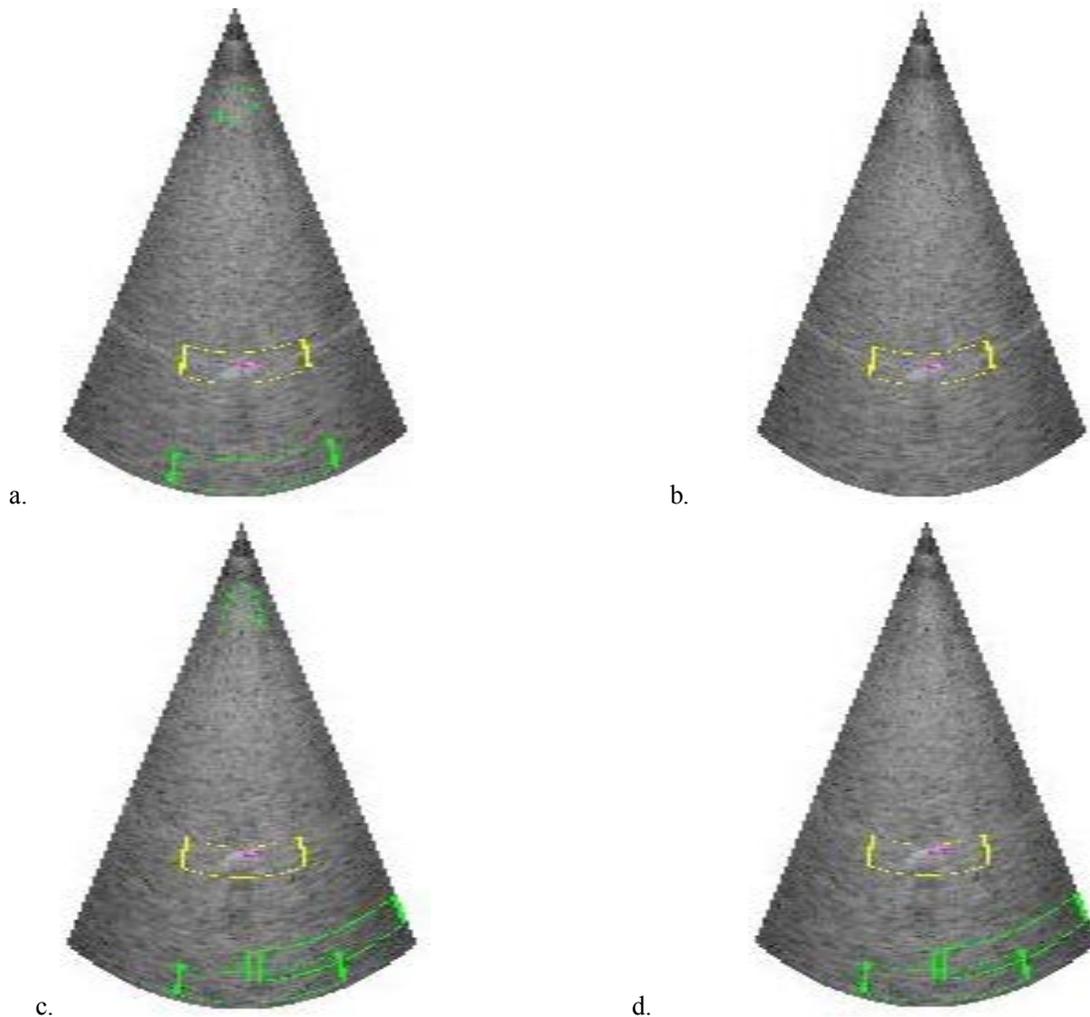


Figure 6. SONAR image with all ROIs labeled (a, c), then eliminated (b, d). The pink bounding box is the true target, green bounding boxes are the ROIs detected, and the yellow bounding box is the ROI containing the true target. In (d), a few green ROIs were missed by the false positive eliminator.

IV. Results

4.1. FROC Curve

The primary indicator of performance used was the Free-Response Receiver Operating Characteristic (FROC) curve. This curve illustrates the trade-off between detection rate (accuracy, Y axis) and false positive rate (X axis). A reliable system aims to converge to or approach its maximum accuracy while staying at or under one or two false positives per image.

After adjusting the filter for each range of SONAR by varying variance(σ), scale and shift, detection was dramatically improved—from an average 20% catch rate using one Mexican-Hat wavelet filter to around 70% using the composite filters. Although the histogram filter was described earlier as a method for normalizing intensity values, it was also found that this filter enhanced noise, making the filters less effective at detecting ROIs. Test one results are displayed below in Figure 7.

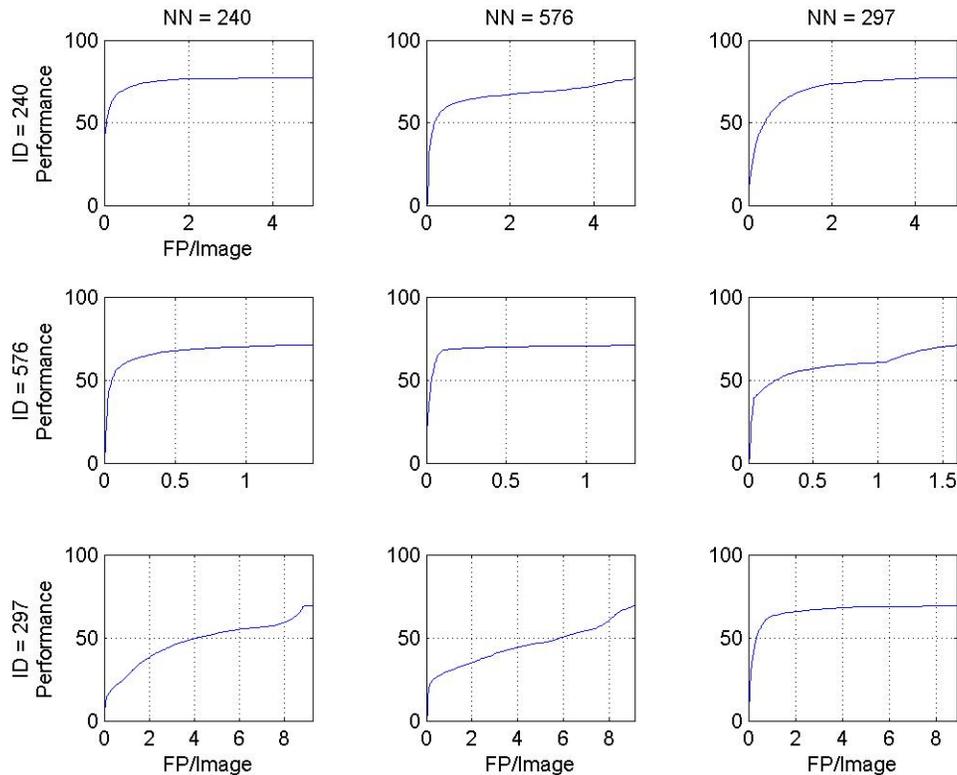


Figure 7. FROC Curve results for three different neural networks. Each column represents the performance of one optimized neural network, while each row is the results from one video. As expected the optimized neural network performs best for its corresponding video, reaching the upper limit of its detection rate while keeping the false positive count to a minimum. The upper limit of performance is determined by the parameters of the filters, which were determined arbitrarily.

For the different training schemes for neural networks, it was to be expected that the neural network would perform best when tested with the same video it was optimized for. However, generally these networks did not perform well for other videos. The third video, ID number 297, was particularly difficult because the target shapes and intensities were extremely different from the other two videos. In order to increase detection rate in this particular video, detection in the other two videos (576 and 240) was decreased.

The mixed neural network performed fairly well across all three videos. Although it did not reach the accuracy and false positive count of any optimized neural network, it also did not have a significant performance drop when tested with any other video. Figure 8 shows the results from test two.

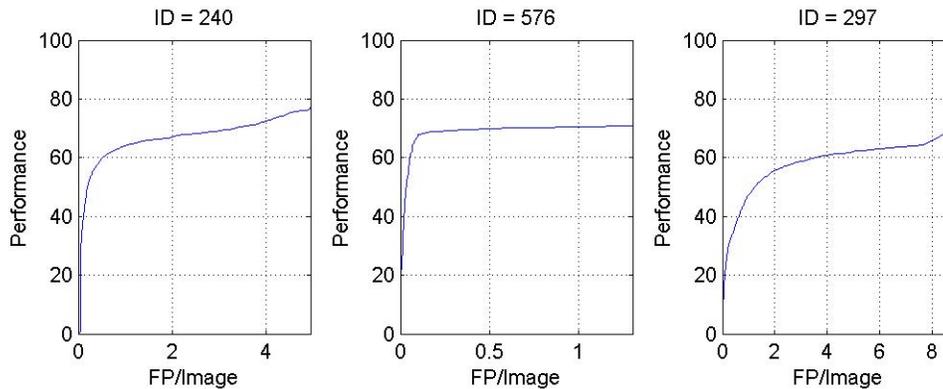


Figure 8. FROC Curve results for the mixed neural network. While performance is slightly lower than each optimized neural network, it does not have a significant performance drop across the three videos.

V. Discussion

Our results support the intuition that it is advantageous to explore different mixed training sets, in order to make an ATR system that is robust across videos. However, there are several other factors that could explain the limited performance of the "optimized" neural networks. First, the problem of overfitting is always a concern. It is possible that the optimized neural networks did not classify well for the other videos because too many training examples were provided. However, due to the disparity between targets across videos, it does not seem likely that even a sparsely trained neural network would be able to make the generalization.

However, the origins of these videos were not specified, and it was unclear as to why the targets varied so dramatically in appearance across videos. It is possible that these videos were taken by different devices. If so, it would be possible to optimize an ATR system for each device, as was done for each video. While such a system would not be able to perform across systems, it could be calibrated to perform highly for a specific device. In this case, the optimized approach would be desirable.

Finally, the decisions on filter parameters, features extracted, training sizes and training set were more or less arbitrary. As previously stated, the filter used to detect ROIs is essential to the maximum detection rate of ATR systems using this framework. Future work could be done to further optimize these filters in order to boost the detection rate of this system. The training process and input features for the false positive eliminator could also be revisited and optimized, to balance the robustness of the classifier with the problem of overfitting.

Acknowledgments

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, and was sponsored by the Undergraduate Student Research Project (USRP), the Department of Defense (DoD) and the National Aeronautics and Space Administration (NASA). I would first like to thank my mentor, Dr. Thomas Lu, for his expert supervision, guidance and support that made completing this project possible. Also, thanks to Dr. Ethan Gallogly and Matthew Scholten for their insights when I was having difficulties with my model.

References

1. Lu, T., Hughlett, C., Zhou, H., Chao, T., and Hanan, J., "Neural network post-processing of grayscale optical correlator," *SPIE Conference* **5908** (2005).
2. Johnson, O., Edens, W., Lu, T., and Chao, T., "Optimization of OT-MACH Filter Generation for Target Recognition," *SPIE Conference* **7340** (2009).
3. Suiter, H., and Wolff, M., "Image preparation for Enhancement of Recorded Underwater Video" *SPIE Conference* **7303**. (2009)

4. Ye, D., Edens, W., Lu, T., and Chao, T., “Neural Network Target Identification System for False Alarm Reduction,” *SPIE Conference* **7340** (2009).
5. Zhang, Y. and Lu, T., “Testing of Haar-like Feature in Region of Interest detection for Automated Target Recognition (ATR) System.” (2010).
6. Zhou, H., Hughlett, C., Hanan, J., and Chao, T., “On the Development of Filter Management Module for Grayscale Optical Correlator,” *SPIE Conference* **5437**, 87–94 (2004).