Convolutional Neural Network based Resolution Enhancement of Underwater Sonar Image Without Losing Working Range of Sonar Sensors

Minsung Sung, Hangil Joe, Juhwan Kim and Son-Cheol Yu
Department of Creative IT Engineering
Pohang University of Science and Technology(POSTECH)
Pohang, Republic of Korea
Email : {kyjor23, roboticist, robot_juhwan, sncyu}@postech.ac.kr

Abstract—**In underwater environment, sonar sensors have the advantage of being able to shoot images in turbid environment and having long working range. However, images taken with sonar sensor are difficult to recognize because of their low resolution. This paper proposes neural network based efficient resolution enhancement method in sonar images. We built convolutional neural network composed of 23 convolutional layers and 18 ResNet blocks, and trained the network with actual and denoised underwater sonar images. As a result, high resolution images can be restored from manually lowered resolution images, recording higher PSNR compared to interpolation algorithms. The proposed method can increase resolution of noisy, low-resolution sonar images without loss in working range.**

Keywords—sonar image enhancement, super resolution, neural network

I. Introduction

Underwater sonar sensors are important for a variety of underwater tasks. Sonar sensors are widely used in underwater tasks because they can be used in a turbid stream and have long working range. Sonar sensors, however, also have disadvantages that the resulting images have severe noise and low resolution. Therefore, images taken with sonar sensors are not intuitive and hard to recognize.

For the more precise task, resolution of sonar image should be increased. One solution to increase resolution is to use a higher frequency acoustic wave when taking the image. Since the resolution of the sonar images is associated with the frequency of the sonar, the higher frequency gives higher resolution. However, the working range of sonar is inversely proportional to frequency due to acoustic nature in underwater [1]. So using a higher frequency to take images will reduce the working range. However, some underwater missions, such as searching sunken ships or airplanes, have limitation to approach targets and the resolution of the sonar images is also restricted. Scanning and searching large area near the seabed is inefficient and nearly impossible. Therefore, it is important to develop a method to increase the resolution of sonar images without increasing frequency of acoustic waves. In this paper, we proposed a neural network based method to increase the resolution of the sonar images.

In section II, we provided other research about increasing resolution of images and enhancing underwater sonar images. In section III, we discussed image set we acquired to train networks, the neural networks that we built, and the way to train the network to increase the resolution of underwater sonar images. In section IV, we provided a result of the network and compared proposed method with other methods based on interpolation by computing peak signal to noise ratio (PSNR). In section V, we summarized our research.

II. Related Work

Many studies have proposed a method for increasing the resolution of sonar images. One approach to increasing the resolution is mosaicing of images. When acquiring sonar images of a certain area, it is possible to obtain higher resolution images of the area by dividing the area, taking multiple images, and then mosaicing multiple captured sonar images. Kim, *et al.* [2] proposed an image fusion algorithm to construct higher resolution images through registration and mosaicing process. In addition, Natalia, *et al.* [3] presented a method to increase the resolution of sonar image by mosaicing the overlapping images by taking the same area through various trajectories of an underwater vehicle. However, these methods have limitations in that resulting images have non-uniform resolution and images taken in short distance are required.

Another approach to increasing the resolution of sonar images is to reduce speckle noise. Speckle noise is the typical noise of underwater sonar sensors generated by the ocean reverberation of acoustic waves. Rithu, *et al.* [4] and Huo, *et al.* [5] proposed a despeckling method to increase the resolution of sonar images to address that speckle noise is a major factor to degrade the resolution of underwater sonar images. Rithu, *et al.* used a Kalman filter to eliminate noise from pixels with speckle noise and estimate the original pixel values. Huo, *et al.* developed a fast method to increase the resolution with despeckling with Curvelet transform. However, there is a limit to the resolution that can be increased with these methods.

978-1-5386-1654-3/18/$31.00 ©2018 IEEE

Some studies have introduced learning-based methods to increase the resolution of sonar images. Liyong, *et al.* [6] developed learning-based super-resolution algorithm for sidescan sonar images. Their method trained two dictionaries composed of low-resolution image patches and high-resolution image patches. After training, super-resolution images are constructed using sparse representation given low-resolution images.

We also proposed learning-based sonar resolution enhancement methods. Using the convolutional neural network, proposed method can increase the resolution of given sonar images up to several times without losing working range and approaching seabed or target object.

III. Proposed Method

A. Super Resolution Neural Network

The classical approach to increase the resolution of images is using interpolation, such as polynomial interpolation and spline interpolation. However, interpolation-based resolution increase methods have limitation. The resulting image has a blurred edge and details are lost.

In particular, for underwater sonar images, methods of increasing the resolution by mosaicing multiple images or increasing the resolution through signal processing have been introduced. However, these methods have the disadvantages that the working range of sonar sensor is reduced or there is a limitation in increasing the resolution.

So, we proposed a neural network-based method to increase the resolution of sonar images by several times without loss of sonar sensor working range and image detail. In recent years, neural networks have been introduced to increase the resolution of images by focusing on image detail restoration.

We adopted super-resolution network model developed by Dahl, *et al.* [7]. The network is composed of two convolutional neural networks (CNN), the prior network and the conditioning network. The prior network consists of 23 convolutional layers and predicts details of the reconstructed high-resolution images. In addition, conditioning network consists of 18 ResNet Blocks. This network predicts each pixel value of high-resolution images to be reconstructed from given low-resolution images.

We implemented and modified the network model and then used the custom underwater sonar image data-set to train the network. The loss function is defined as cross-entropy between ground truth high-resolution image and the predicted high-resolution images. By training the network to minimize loss, high-resolution images could be predicted when given low-resolution images.

B. Data Acquisition

To train the network, we acquired actual underwater sonar images. Increasing resolution using the neural network is to find the relation between low-resolution and high-resolution images. In order for the network to find relationship and the weights of the networks to converge, similar geometry among the images is required. So we decided to use the images of one object to train the network.

TABLE I. Specification of Acoustic Lens Based Image Sonar Used In Data Acquisition

Frequency	Working range	Field of View
1.1 MHz	30 m	29 ˚

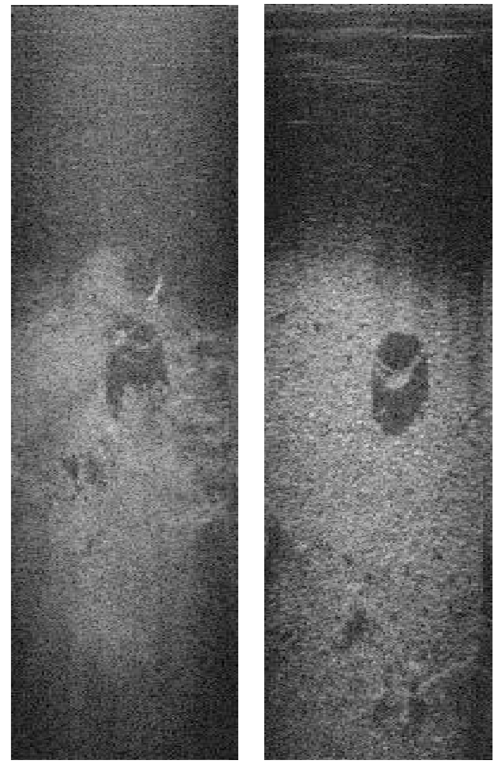


Fig. 1. Samples of acquired image of two types of tires

The tires were put into the seabed and images were taken with acoustic lens-based image sonar named Dual Frequency Identification Sonar (DIDSON). The specifications of the sonar sensor we used are described in Table I. Furthermore, we thought that using only one type of object would make the network overfit. So we used images of two types of tires. Fig. 1 is samples of images acquired to train the network.

A total of 1,057 images of tires were acquired. The image set is divided into two groups: 957 images for training and 100 images for testing.

C. Training the Network

The original size of the acquired sonar images is 512 by 96 pixels. To reduce computation and correlate training image sets to train the network efficiently, we cropped only the tire region from entire images and resized it to 32 by 32 pixels. Then we trained the network using 32 by 32 images and corresponding images manually downsampled to 8 by 8 pixels.

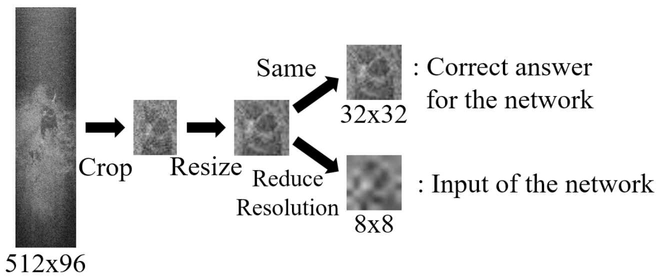
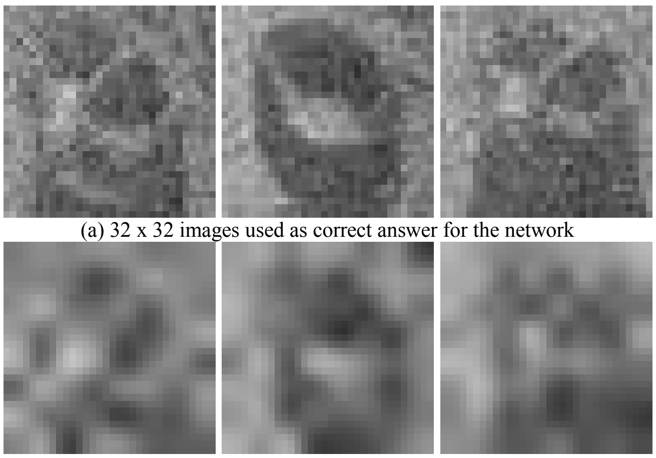


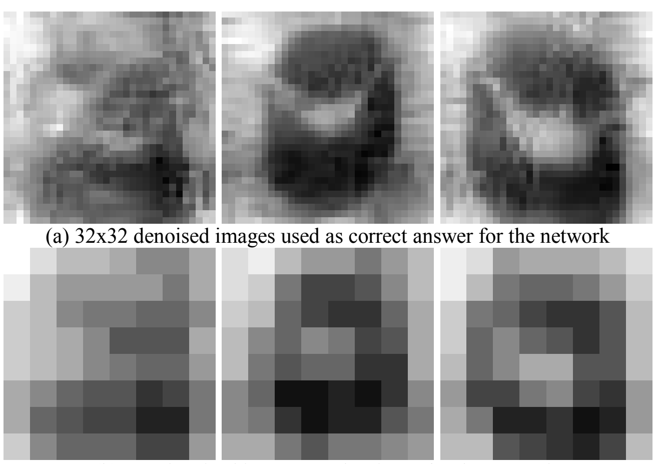
Fig. 2. Process to make a pair of training images



(a) 32 x 32 images used as correct answer for the network

(b) corresponding 8x8 images used as input for the network

Fig. 3. Samples of training image set



(a) 32x32 denoised images used as correct answer for the network

(b) 8x8 denoised images used as input for the network.

Fig. 4. Samples of denoised training image set

TABLE II. HYPER-PARAMETERS USED TO TRAIN THE NETWORK

Optimizer	**Batch size**	**Number of epoch**	**Learning rate**	**Activation function**
RMSProp Optimizer	16	162,000	0.0004	tanh

The network could be trained to restore a 32 by 32 image from given 8 by 8 image using an 8 by 8 image as input and a 32 by 32 image as the correct answer. Fig. 2 illustrates the process of creating high-resolution and low-resolution image pairs to train the network. In addition, Fig. 3 shows samples of images used to train the network.

Furthermore, severe noise from underwater sonar images could interfere with network convergence and degrade the quality of resulting images. So we use wavelet transform to add denoised images to train the network. Fig. 4 shows samples of denoised image pairs used to train the network.

Finally, hyper-parameters should be determined appropriately to train the network efficiently and make the network predict output well. A grid search was conducted changing the value of hyper-parameters. As a result, hyper-parameters maximizing the quality of resulting images from the network were chosen like Table II.

IV. TRAINING EXPERIMENT

We conducted experiments to determine the parameters needed to train the network well. Training and testing were done in an environment using Graphics Processing Units (GPU) Titan X.

For the training image set, we tried to use images of various sizes from 512 by 96 pixels to 32 by 32 pixels. When using an image larger than a 32 by 32 pixels image, each training epoch took too much time to conduct. In addition, there were too many weights and biases in the network, making it difficult to get the network to converge. Finally, when using large images, the resulting image was unnatural and discontinuous because of a large number of noise pixels in the image.

We also trained the network by adding denoised images to the training dataset. Noise in the sonar image degraded the quality of the network and made the resulting image look unnatural and discontinuous. However, if too many denoised images were added to the training dataset, the quality of the network was also degraded. Because the denoised image has different pixel values than the original sonar image, the network could not find the relationship among training dataset and did not learn the weights and biases well. Finally, 100 denoised images were added to the training image set, resulting in a total of 1,057 images used to train the network.

Finally, a grid search to find optimal hyper-parameters for the network were conducted. Cross-entropy between the ground truth and the predicted image was measured while changing the value of the batch size, learning rate, optimizer, and the number of epochs. As a result, a combination of parameters that minimize cross-entropy was selected as shown in Table II and was used for prediction of high-resolution images.

V. RESULT

A. *Training Result*

The network was trained by 162,000 epochs and took 18 hours. For each training epoch, 16 images were randomly selected

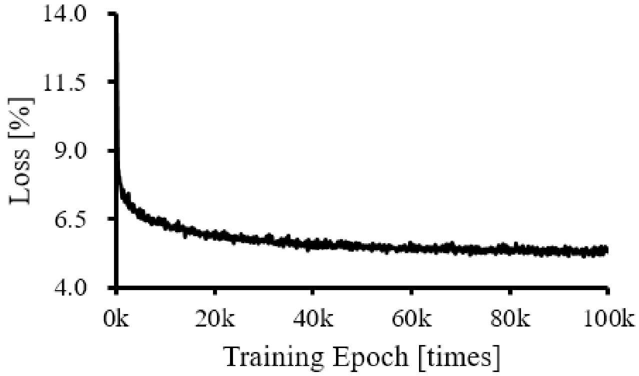
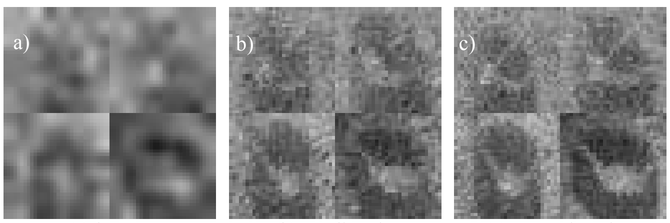
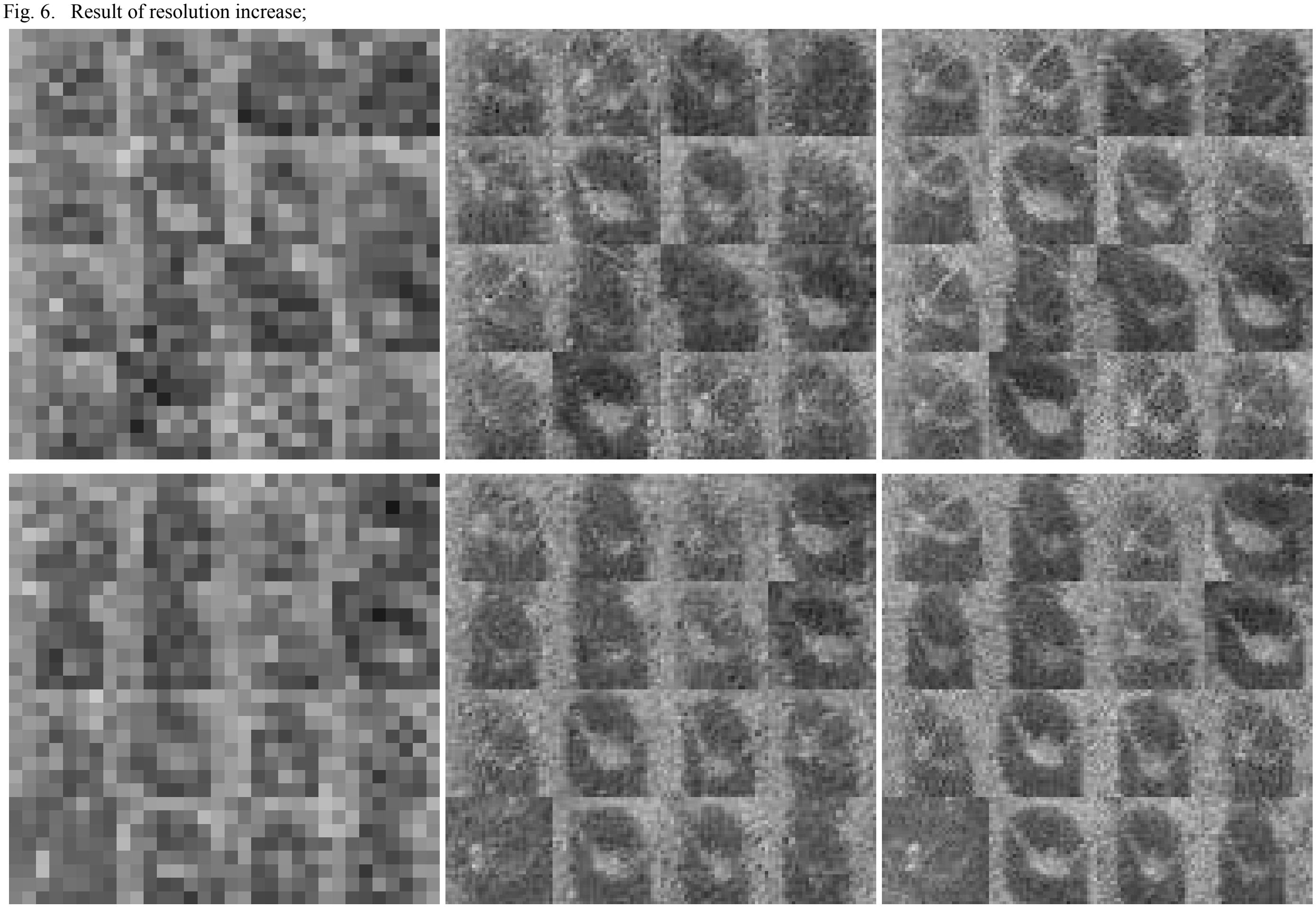


Fig. 5. The loss value function of training epoch



(a) input (b) predicted (c) ground truth

Fig. 6. Result of resolution increase;



(a) input low-resolution images (b) predicted high-resolution images (c) ground truth images

Fig. 7. Samples of output from the network

from the learning images set to train the network. The loss value, defined as the cross-entropy between the predicted high-resolution image and the ground truth images, decreased as the training progressed, eventually dropping to 5.12 %. Fig. 5 shows the loss value converges as the training progresses.

B. Resolution Enhancement Result

After network training was completed, the network tested with 100 sonar images taken of tires. Fig. 6 shows the result of the network. Fig. 6a is images of manually reducing the resolution to 8 by 8 pixels. From these images, the network predicted 32 by 32 pixels images as shown in Fig. 6b.

Compared to the ground truth image (Fig. 6c), the network shows that it is possible to increase the resolution to the original resolution while restoring details of the image such as shadows, highlights and object outlines. Additional results of the network to increase resolution are shown in Fig. 7.

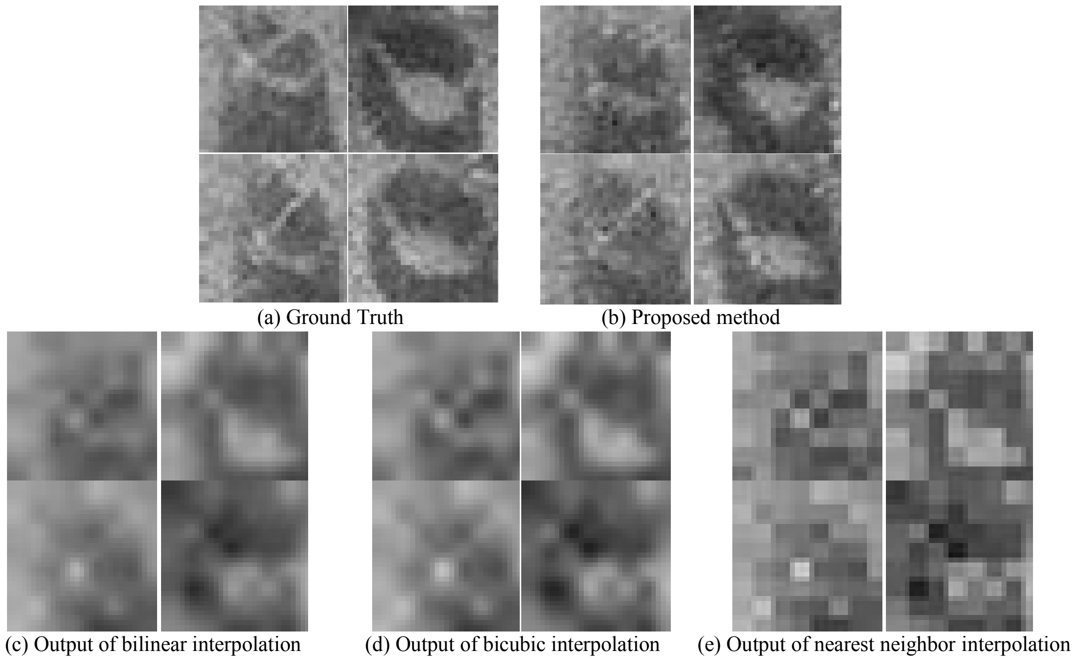


Fig. 8. Samples of output from proposed method and interpolation-based methods

TABLE III. COMPARISON AMONG PROPOSED METHOD AND INTERPOLATION-BASED ALGORITHMS ABOUT PSNR

Interpolation			**Proposed Method**
Nearest Neighbor	***Bilinear***	***Bicubic***	
45.65	48.89	47.73	51.23

C. Comparison to other methods

To analyze the quality of the proposed method, we compared the output of proposed method with interpolation-based resolution increase methods; Nearest Neighbor (NN), Bilinear, and Bicubic interpolation. In Fig. 8, the results of proposed method represent the details better than the interpolation-based methods.

We also calculated the peak signal to noise ratio (PSNR) of the resulting images for quantitative measurement of the output. PSNR indicates how closely the manually lowered resolution image is restored to the original image. As shown in Table III, proposed method recorded a higher PSNR than interpolation-based methods.

VI. CONCLUSION

In this paper, neural network-based resolution enhancement method for underwater sonar images is proposed. We built CNN and trained the network using actual underwater sonar images. As a result, images can be restored from manually lowered resolution to resolution of original images.

In addition, the proposed method provided better image detail and higher PSNR than the interpolation-based methods. The proposed method is important because it does not lose working range of sonar sensors and it can provide high-resolution underwater sonar images without reaching the target in underwater missions.

ACKNOWLEDGMENT

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ICT Consilience Creative program(IITP-2017-R0346-16-1007) supervised by the IITP(Institute for Information & communications Technology Promotion).

This research was a part of the project titled 'Gyeongbuk Sea Grant', funded by the Ministry of Oceans and Fisheries, Korea.

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2017R1A5A1014883).

REFERENCES

[1] H. R. Gordon. "Can the Lambert-Beer law be applied to the diffuse attenuation coefficient of ocean water?." Limnology and Oceanography 34.8, 1989, pp.1389-1409

[2] Kim, K., N. Neretti, and N. Intrator. "Acoustic camera image mosaicing and super-resolution." *OCEANS'04. MTTS/IEEE TECHNO-OCEAN'04*. Vol. 2. IEEE, 2004.

[3] Hurtós, Natalia, Yvan Petillot, and Joaquim Salvi. "Fourier-based registrations for two-dimensional forward-looking sonar image

mosaicing." *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*. Ieee, 2012.

[4] James, Rithu, and M. H. Supriya. "Spatial resolution enhancement of sonar images using one step pixel prediction." *Control Communication & Computing India (ICCC), 2015 International Conference on*. IEEE, 2015.

[5] Guanying, Huo, Li Qingwu, and Fan Xinnan. "A fast super-resolution algorithm with despeckling for multi-frame sonar images." *Information Science and Engineering (ICISE), 2010 2nd International Conference on*. IEEE, 2010.

[6] Ma, Li Yong, Xi Li He, and Zi Shen Huang. "Sidescan sonar image super resolution based on sparse representation." *Applied Mechanics and Materials*. Vol. 48. Trans Tech Publications, 2011.

[7] Dahl, Ryan, Mohammad Norouzi, and Jonathon Shlens, "Pixel Recursive Super Resolution," arXiv preprint arXiv:1702.00783 (2017).