Opti-Acoustic Reciprocal Feature Matching

Anuj Zore
M.Eng. in Robotics
University of Maryland, College Park
zoreanuj@umd.edu

Amogh Wyawahare
M.Eng. in Robotics
University of Maryland, College Park
wamogh@umd.edu

***Abstract*—In this study, we introduce a robust method for domain adaptation between Sonar and RGB modalities, aiming to significantly enhance the interoperability and modality transfer capabilities of sonar images. By enabling effective translation between these modalities, our approach facilitates improved accuracy in feature detection, motion estimation, and other critical tasks. We conducted comprehensive evaluations using two transfer approaches: a) Sonar to RGB and b) RGB to Sonar, employing state-of-the-art models such as pix2pix and pix2pixHD. Our experimental results showcase the method's ability to recover detailed spatial information from sonar data, even in conditions where RGB data is severely distorted. These promising outcomes underscore the potential of our approach in overcoming the limitations of individual modalities, thereby improving the accuracy and reliability of applications in challenging environments.**

I. Introduction

Marine exploration, including underwater inspections and maintenance, plays a crucial role in various applications, from scientific research to industrial operations. These activities are inherently hazardous, with safety concerns amplified in complex environments such as ports, piers, industrial basins, and channels. Navigating and self-localizing in these settings can be challenging due to poor visibility, which often worsens as activities stir up sediments, leading to white-out situations where optical monitoring becomes ineffective.

Sonar imaging provides a viable alternative to visual sensing under such conditions, as it remains unaffected by poor visibility and lighting. However, interpreting sonar images is difficult due to their low signal-to-noise ratios. In our work, we introduce an advanced method for domain adaptation between Sonar and RGB modalities, aiming to enhance the interoperability and usability of sonar images. Our approach facilitates accurate feature detection, motion estimation, and other tasks by translating sonar images into more intuitive visual-like representations.

We evaluated two transfer methods—Sonar to RGB and RGB to Sonar—using state-of-the-art models like pix2pix and pix2pixHD. Our results demonstrate the efficacy of our approach in recovering spatial information from sonar data, even in conditions where RGB data is heavily distorted. This technique not only improves the accuracy of marine exploration tasks but also ensures that human operators can effectively interpret underwater scenes, thereby enhancing the safety and efficiency of operations in turbid or dark waters.

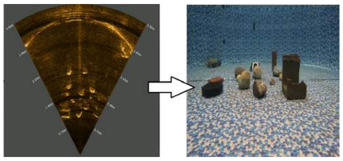

Fig. 1: Case without Turbidity

Fig. 2: Case with High Turbidity

A. Related Work

Recent advancements in deep learning have significantly enhanced the translation of sonar images to RGB images, providing clearer and more intuitive visual representations. Techniques such as Conditional GANs [2] (CGANs) have been pivotal, with frameworks like pix2pix and its high-definition variant, pix2pixHD [3] [4], demonstrating success in this domain. Pix2pix leverages CGANs combined with U-Net architecture and PatchGAN to maintain image consistency and capture high-frequency details. Pix2pixHD extends this approach by introducing a coarse-to-fine generator and multi-scale discriminators, enabling the generation of high-resolution images. This method ensures that the translated images are photorealistic and temporally consistent, even for continuous streams of sonar and optical data.

One notable application is fish monitoring, where deep learning models translate low-quality sonar images into realistic daytime RGB images. This capability is especially beneficial in low-visibility conditions, such as nighttime or turbid waters, where optical cameras fail. By integrating sonar and optical data, models like the Nightvision algorithm successfully generate interpretable visual images from sonar inputs.

The primary challenge in sonar to RGB image translation lies in the substantial differences between sonar and optical imaging modalities. Sonar images, based on acoustic signals, have a different geometric representation compared to optical

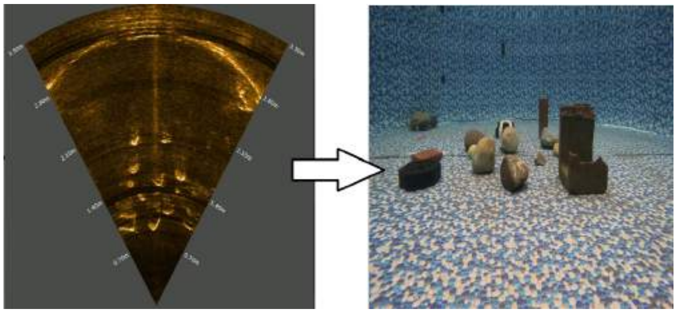

Fig. 3: Overall Task of Domain Adaptation

images, which rely on light. This disparity makes direct conversion a complex and ill-posed problem. Additionally, the scarcity of paired sonar and RGB datasets poses a significant hurdle for training effective neural network models.

Despite these challenges, the advancements in deep learning methodologies for sonar to RGB image translation have vast practical applications. Improved image translation facilitates better inspection and maintenance of underwater structures, enhances search and rescue operations, and improves environmental monitoring. By converting sonar data into familiar RGB formats, operators can make more informed decisions, benefiting various underwater applications and enhancing the overall safety and efficiency of marine operations. Future research should focus on addressing data scarcity and refining models to improve translation accuracy and reliability across diverse underwater environments.

II. Methodology

In this work, we employ Pix2pixHD for translating sonar images to RGB images, enhancing underwater visual interpretation. The pipeline begins with video frames provided to us, which are then preprocessed through cropping, augmentation, and denoising. These preprocessed frames are loaded into a dataloader.

Later, during inference images are translated from sonar to its approximate rgb counterpart as shown in Figure 3.

A. *Fundamental Engineering Principles*

In this project, we applied the fundamental engineering principle of abstraction in the design of our deep learning models, particularly critical for our domain adaptation task between sonar and RGB imaging. Abstraction in engineering involves reducing complexity by focusing on the high-level structure without being bogged down by details, which we implemented by abstracting the complex behaviors of sonar and RGB data into more manageable representations through deep learning models like pix2pix and pix2pixHD. This principle facilitated the modeling of intricate and noisy underwater data, allowing our neural networks to focus on essential features without getting overwhelmed by irrelevant data noise. Utilizing abstraction, our models could effectively learn and generalize the critical characteristics of the data, translating sonar inputs into visually intuitive RGB outputs. This approach not only streamlined our model training process but also improved the robustness and accuracy of our system, demonstrating a direct application of engineering principles to overcome challenges in deep learning tasks.

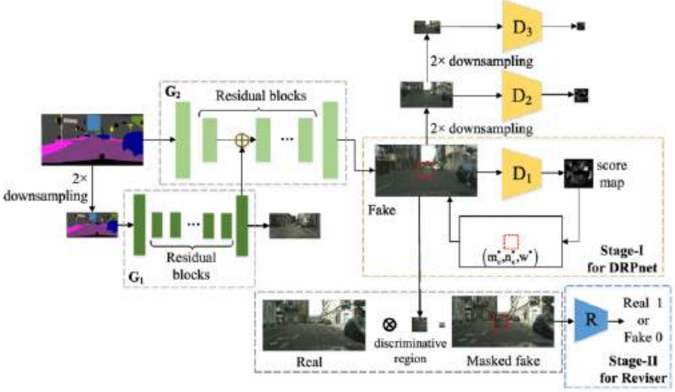

Fig. 4: Pix2Pix HD Architecture

B. *Pix2pixHD Architecture*

Pix2pixHD extends the capabilities of the original pix2pix framework by introducing a coarse-to-fine generator and multi-scale discriminators. The generator is divided into a global generator network and a local enhancer network. The global generator captures the overall structure of the image, while the local enhancer refines high-frequency details. Multi-scale discriminators, operating at different image scales, ensure the generation of realistic high-definition images.

1) Encoder-Decoder Structure: The generator uses an encoder-decoder structure with skip connections (U-Net architecture) to retain fine details. The encoder comprises several downsampling convolutional layers, while the decoder consists of upsampling layers. Each layer uses Convolution-BatchNorm-ReLU (CBR) blocks to maintain image quality.

2) Training: The network is trained using a CGAN framework, where the generator aims to produce realistic images while the discriminator attempts to distinguish between real and generated images. The training process involves minimizing both adversarial loss and reconstruction loss to ensure image fidelity and realism. Dropout layers are added to the decoder to prevent overfitting. he model is trained for 100 epochs with a batch size of 8, using the Adam optimizer with a learning rate of 0.0002 and momentum parameters.

3) Detailed Pipeline Steps:

- Video Frames: The process starts with video frames provided by the professor.
- Preprocessing: The frames undergo preprocessing, which includes cropping to focus on relevant areas, augmentation to create varied training data, and denoising to enhance image quality.
- Data Loader: The preprocessed images are then fed into a dataloader, which prepares the data for training.
- Pix2pixHD Implementation: Pix2pixHD is employed for the image-to-image translation task. The architecture

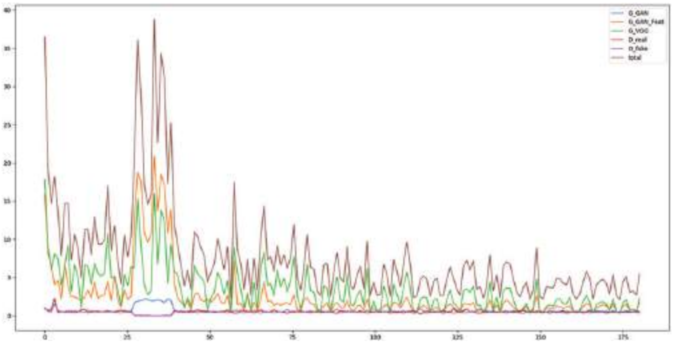

Fig. 5: Pix2PixHD training Loss

leverages a coarse-to-fine generator to handle high-resolution image generation and multi-scale discriminators to maintain realism across different scales.

- Training: The CGAN framework is used for training, where the generator and discriminator are trained simultaneously. The generator produces RGB images from sonar inputs, while the discriminator distinguishes between real and generated images.
- Optimization: The model is optimized using the Adam optimizer, with careful tuning of learning rates and momentum parameters to ensure effective training.
- Evaluation: The Structural Similarity (SSIM) index is used to evaluate the quality of the translated images, providing a measure of how similar the generated images are to real RGB images.

4) Evaluation Metrics: The Structural Similarity (SSIM) index is used as an evaluation metric, which measures the perceptual similarity between images. SSIM values range from 0 (dissimilar) to 1 (similar), allowing robust assessment under noise and distortion influences.

By leveraging deep learning techniques, particularly Pix2pixHD, this work effectively translates sonar images into more intuitive and visually appealing RGB images. This approach significantly improves underwater visualization and interpretation, benefiting various applications such as underwater inspections, environmental monitoring, and marine research. Future research should focus on addressing data scarcity and refining models to improve translation accuracy and reliability across diverse underwater environments.

III. Results

In this work, we employ Pix2pixHD to translate sonar images into RGB images, enhancing underwater visual interpretation. The pipeline starts with video frames provided by a professor, which undergo preprocessing steps such as cropping, augmentation, and denoising to enhance image quality. These preprocessed frames are loaded into a dataloader for further processing. Pix2pixHD, an advanced image-to-image translation model, is used for this task. The architecture includes a coarse-to-fine generator, which captures overall image structure and refines high-frequency details, and multi-scale discriminators that ensure realistic high-definition images. The encoder-decoder structure with skip connections (U-Net architecture) helps retain fine details, while Convolution-BatchNorm-ReLU (CBR) blocks maintain image quality. The model is trained using a Conditional GAN (CGAN) framework, where the generator produces realistic RGB images from sonar inputs and the discriminator distinguishes between real and generated images.

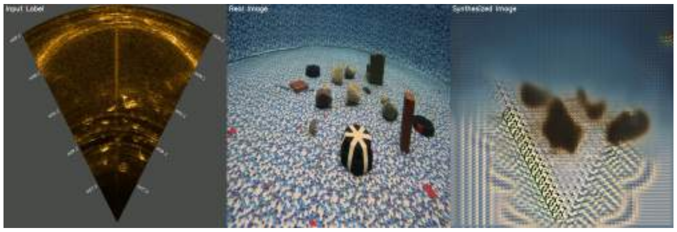

Fig. 6: Epoch 1

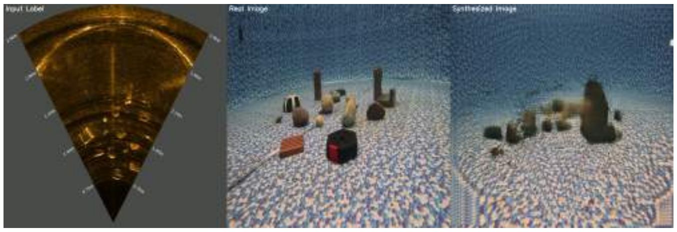

Fig. 7: Epoch 5

Training involves minimizing both adversarial and reconstruction losses to ensure fidelity and realism. The model is optimized with the Adam optimizer for 40 epochs, with a learning rate of 0.0002 and momentum parameters 1 = 0.5 and 2 = 0.999. The VGG inception loss, which evaluates output quality, dropped to 0.12, indicating high sensitivity to noise and artifacts.

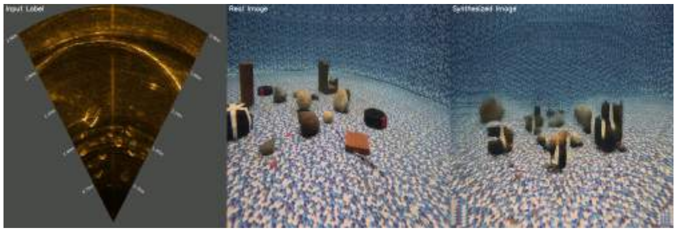

Fig. 8: Epoch 10

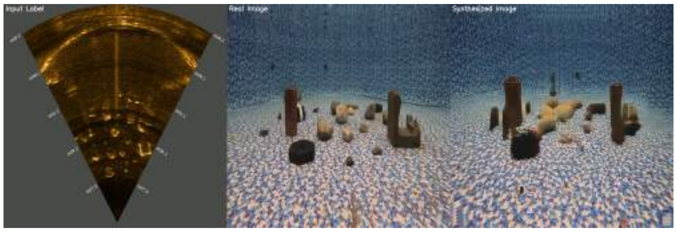

Fig. 9: Epoch 15

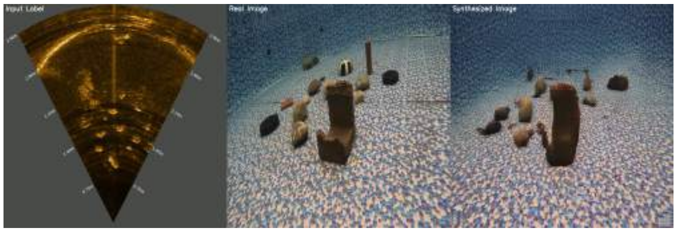

Fig. 10: Epoch 20

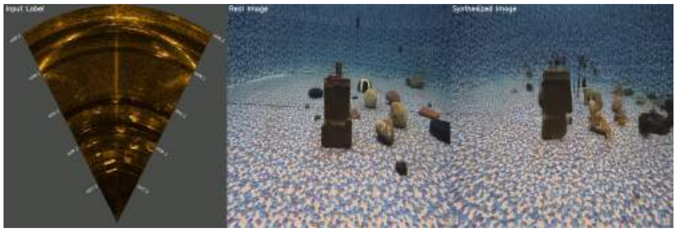

Fig. 11: Epoch 25

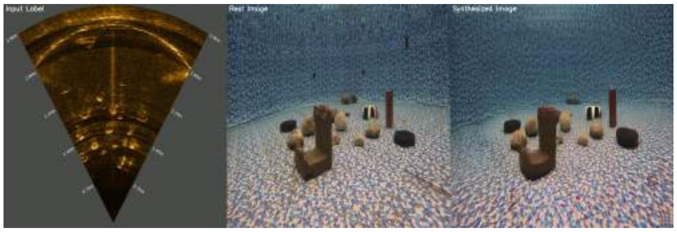

Fig. 12: Epoch 30

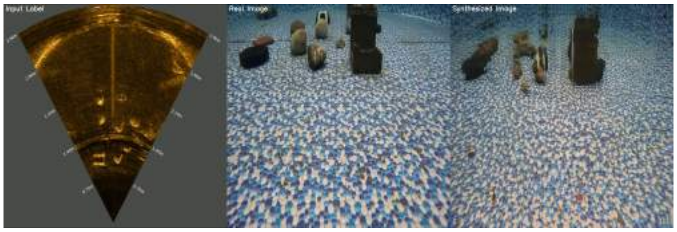

Fig. 13: Epoch 35

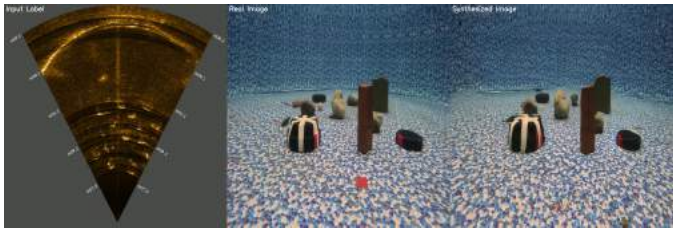

Fig. 14: Epoch 40

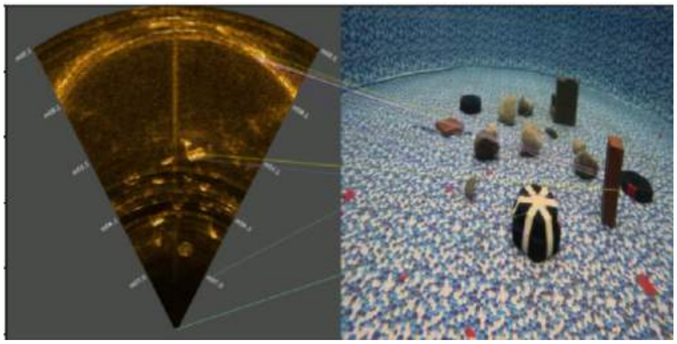

Fig. 15: Sonar and RGB Feature Matching - 2/14 Matches

A. *Quality Evaluation*

From Figure 15 to 20, we observe a notable improvement beginning from the 10th epoch, with the translated images demonstrating increasingly high fidelity to the ground truth. This progression underscores the capabilities of deep learning technologies, particularly the pix2pix and pix2pixHD models, in translating sonar images into more intuitive RGB images. This enhancement significantly improves underwater visualization and interpretation, facilitating critical applications such as underwater inspections, environmental monitoring, and marine research.

The increasing fidelity of the image translation can be quantitatively assessed by evaluating the number of correct feature matches between the generated images and the ground truth RGB images. Initially, at Epoch 1 (Figure 16), the number of correct matches is relatively low, reflecting the model's early stage of learning and its limited ability to capture and replicate the intricate details and patterns of the RGB modality from sonar inputs. However, as the training progresses, the model increasingly adapts to the nuances of the data, evidenced by a substantial increase in correct matches by Epoch 40 (Figure 20). This improvement is not just a measure of visual similarity but also an indicator of the model's growing accuracy in representing and interpreting complex underwater features and conditions.

The deep learning approach employed leverages a combination of adversarial training and feature consistency, enabling the model to refine its outputs iteratively. As such, future research should focus on addressing challenges such as data scarcity, which limits the model's exposure to diverse environmental conditions. Refining these models further could enhance their translation accuracy and reliability, allowing for broader application across various and variable underwater environments. This will involve not only expanding the datasets used for training but also incorporating more robust mechanisms for generalization to ensure that the models perform well in untrained settings, ultimately leading to more reliable and accurate systems for real-world applications.

IV. Conclusion

Our work has demonstrated the significant potential of domain adaptation, particularly through the application of sonar-

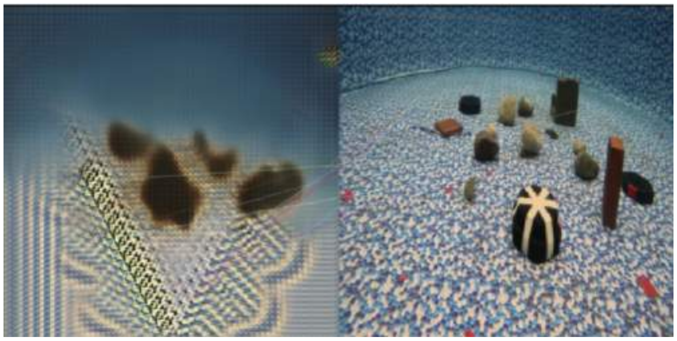

Fig. 16: Generated Image and RGB Feature Matching (Epoch 1) - 15/43 Matches

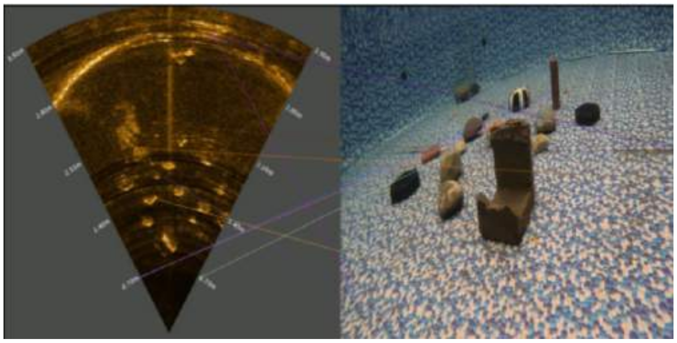

Fig. 17: Sonar and RGB Feature Matching - 3/16 Matches

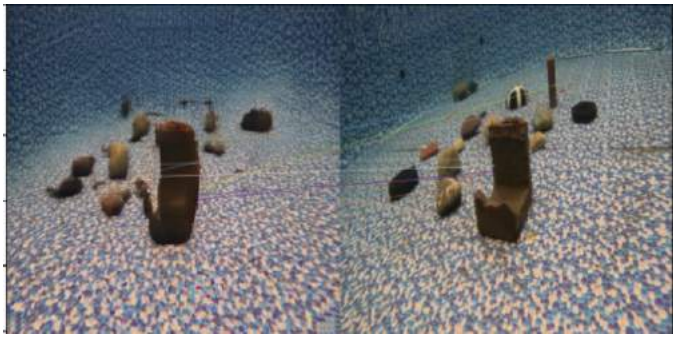

Fig. 18: Generated Image and RGB Feature Matching (Epoch 20) - 32/43 Matches

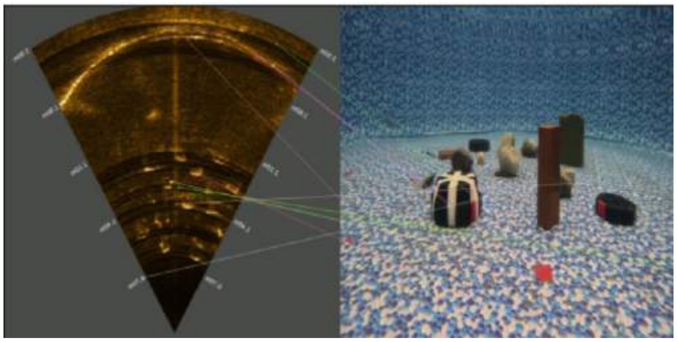

Fig. 19: Sonar and RGB Feature Matching - 2/15 Matches

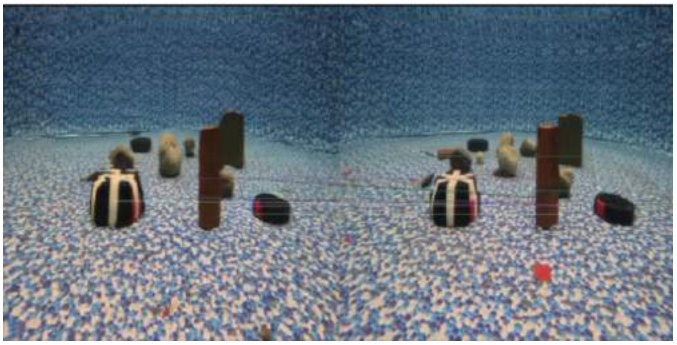

Fig. 20: Generated Image and RGB Feature Matching (Epoch 40) - 41/43 Matches

RGB image pairs for feature matching, motion estimation, and scenario comparison tasks. The conversion of sonar images to the optical image style not only produced outputs of high aesthetic quality but also maintained stability in quantitative assessments, affirming the effectiveness of our approach within controlled conditions.

However, this study has also highlighted several limitations that future work needs to address. Firstly, our assumptions that sonar and optical data are collected from identical locations, configurations, and environments may not hold true in practical, diverse field conditions. This presents a gap between experimental settings and real-world applications.

Secondly, while our pipeline performs well under conditions similar to those it was trained on, such as turbidity and illumination levels, it struggles to generalize across variably different environments. This limitation underscores the need for models that adapt more robustly to changes in external conditions.

Lastly, the requirement for additional fine-tuning when applying our method to different sonar datasets indicates a need for more adaptable or universally applicable solutions. Future research should focus on overcoming these drawbacks to enhance the practicality and applicability of domain adaptation techniques in real-world scenarios, potentially through the development of more dynamic models or the incorporation of adaptive learning mechanisms.

By addressing these issues, we can move closer to realizing the full potential of domain adaptation in underwater imaging and beyond, paving the way for broader applications in diverse and challenging environments.

V. Acknowledgment

Both Anuj and Amogh equally contributed on all sections including data election, data preprocessing, model selection, model training and inference.

References

[1] Negahdaripour, S., Sekkati, H., & Pirsiavash, H. (2009). Opti-acoustic stereo imaging: On system calibration and 3-D target reconstruction. IEEE Transactions on image processing, 18(6), 1203-1214.

[2] Qu Y, Chen Y, Huang J, Xie Y. Enhanced pix2pix dehazing network. InProceedings of the IEEE/CVF conference on computer vision and pattern recognition 2019 (pp. 8160-8168).

[3] Basu, Arpan, et al. ”U-Net versus Pix2Pix: a comparative study on degraded document image binarization.” Journal of Electronic Imaging 29.6 (2020): 063019-063019.

[4] Aggarwal, Karan, et al. ”Benchmarking regression methods: A comparison with CGAN.” arXiv preprint arXiv:1905.12868 (2019).

[5] Xu, Wanru, et al. ”Prediction-cgan: Human action prediction with conditional generative adversarial networks.” Proceedings of the 27th ACM international conference on multimedia. 2019.

[6] Negahdaripour, S., Pirsiavash, H., & Sekkati, H. (2007, June). Integration of motion cues in optical and sonar videos for 3-D positioning. In 2007 IEEE Conference on Computer Vision and Pattern Recognition (pp. 1-8). IEEE.

[7] A. Marburg and A. Stewart, ”Extrinsic calibration of an RGB camera to a 3D imaging sonar,” OCEANS 2015 - MTS/IEEE Washington, Washington, DC, USA, 2015, pp. 1-6, doi: 10.23919/OCEANS.2015.7404377. keywords: Cameras;Calibration;Sensors;Three-dimensional displays;Sonar measurements,

[8] Terayama K, Shin K, Mizuno K, Tsuda K. Integration of sonar and optical camera images using deep neural network for fish monitoring. Aquacultural Engineering. 2019 Aug 1;86:102000.

[9] Kim HG, Seo J, Kim SM. Underwater optical-sonar image fusion systems. Sensors. 2022 Nov 3;22(21):8445.

[10] Połap D, Wawrzyniak N, Włodarczyk-Sielicka M. Side-scan sonar analysis using ROI analysis and deep neural networks. IEEE Transactions on Geoscience and Remote Sensing. 2022 Jan 27;60:1-8.