

Review on Underwater Vision Revolution using Deep Learning for Accurate Fish Species Classification

Aishwarya G Danaraddi ^[1], Anamika Tiwari ^[2], Andani Gowda ^[3], Adithya K M ^[4], Chandra Naik ^[5]

Department of Computer Science and Engineering, Alva's Institute of Engineering & Technology, Moodbidri.

Abstract— Underwater fish recognition (UFSR) is important for marine research and requires re-evaluation of the manual process due to its complexity and cost limitations. This article reviews fish detection and classification techniques, focusing on the use of neural networks (CNN) to accurately classify underwater images. At the same time, an automatic fish classification method using CNN, deep learning and image processing is proposed, achieving 96.29% accuracy and improving the discrimination accuracy. This study also provides a brief review of recent applications, including 13 projects on fish acoustic echo classification using sonar and echo sounder. This study demonstrates many activities, structures, products, and challenges that indicate the possibility of spread from fish to other marine organisms.

Index Terms—Category perceptual loss, deep learning, fish recognition, object detection, species classification.

destructive, portable, and cost-effective solutions to obtain

I. INTRODUCTION

Monitoring the behavior of different fish species is crucial for understanding marine ecosystems, providing essential information about ecosystem health and assessing environmental change ^[14]. Visualizing fish distribution aids in tracking their movements and identifying patterns in their activities, contributing to a comprehensive understanding of the species ^[5]. Despite some progress in fish distribution management, significant variations persist in the information obtained from underwater videos due to challenges like noise, distortion, aliasing, segmentation errors and occlusion ^[15]. Traditional methods, such as flash and background subtraction, face limitations such as color distortion, inconsistent lighting, and the presence of debris and underwater vegetation ^[10].

Applications in the marine industry, such as fish monitoring for commercial fishing or ecological research, can benefit from instantaneous fish detection, location, and distribution ^[7]. Fish detection systems are valuable for measuring fish populations, classifying species, and studying migrations ^[8]. Parameterizing fish in fishing operations complements species and distribution estimation, facilitating the quantification of catch and discard costs ^[22]. The proposed method introduces Convolutional Neural Networks (CNNs), providing powerful and flexible solutions for fish classification, especially with large datasets ^[22]. Utilizing information from the Fish4Knowledge project, the classification involves pre-processing techniques like Gaussian blur, morphological functions, Otsu thresholding, and pyramid mean transform before feeding the improved image into the CNN for distribution ^[14]. This automated process proves essential for accurate and precise fish classification, reducing dependence on time-consuming and error-prone manual procedures ^[14].

The increasing need for electronic monitoring, reporting, and intelligent fish detection underscores the demand for non-

high-quality images ^[18]. Machine learning methods, particularly deep learning methods like CNNs, stand out for their performance in image processing, enabling instant fish detection and classification ^[21]. This paper employs a CNN model to identify fish images, contributing to fisheries research and ocean exploration using data collected from research centers and archives ^[22].

II. LITERATURE REVIEW

In recent years, the application of computer vision techniques in the field of fish species recognition and underwater imaging has gained significant traction. This review synthesizes findings from various studies focusing on different aspects of fish species recognition, technologies classification.

Early research by Chen, Mulgrew, and Grant ^[1] explored the use of radial basis function networks for digital communications channel equalization, providing a foundation for subsequent work in pattern recognition. Duncombe ^[2] assessed the feasibility of infrared navigation, laying groundwork for sensing technologies applicable to underwater environments.

Advancements in image processing and watermarking techniques have enabled resilient public watermarking for images ^[3], while computer vision methodologies have been applied to automate measurement of fish species and length ^[4-5], emphasizing the importance of automated systems in fisheries management.

Techniques such as aggregated segmentation ^[7], identification, and counting of live fish ^[8], and robust feature extraction using neural networks ^[9] have further enhanced the accuracy and efficiency of fish species recognition systems. Research by Chuang et al. ^[7] introduced a novel approach for aggregated segmentation of fish from conveyor belt videos,

demonstrating the feasibility of automated fish processing in industrial settings. Castignolles et al. [8] developed methods for the identification and counting of live fish using image analysis, contributing to advancements in aquaculture monitoring and management.

Recent developments in deep learning have revolutionized the field, with studies demonstrating the effectiveness of convolutional neural networks (CNNs) in underwater fish detection [10-12], species identification [13], and activity tracking [18]. Notably, Joly et al. [13] participated in the Lifeclef 2015 challenge, focusing on multimedia life species identification, highlighting the growing interest in interdisciplinary competitions to benchmark fish species recognition algorithms.

Furthermore, researchers have explored the integration of deep learning with other technologies such as LiDAR [26], stereo-vision [32], and human-robot interaction systems [32], extending the capabilities of underwater imaging and analysis. The study by Gomez Chavez et al. [32] introduced the CADDY underwater stereo-vision dataset for human-robot interaction (HRI), providing valuable resources for training and evaluating underwater vision systems in complex environments.

Datasets such as PlanktonSet [28], Fish-4-Knowledge [29], and Fish-gres [30] have played a crucial role in benchmarking and training deep learning models for fish species recognition. Additionally, studies have addressed challenges such as feature selection [34], classification of ocean ecosystems plankton [35], and exploiting cyclic symmetry in CNNs [39].

Overall, the literature highlights the interdisciplinary nature of fish species recognition research, incorporating elements of computer vision, machine learning, and marine biology to address pressing environmental and conservation challenges in aquatic ecosystems. Collaborative efforts between researchers, industry stakeholders, and policymakers are essential to further advance the field and ensure sustainable management of marine resources.

III. RELATED WORKS

Finding fish species in confined habitats is difficult due to complex backgrounds and noise in images. A few years ago, scientists worked on the same project to use better methods to classify fish species in their natural habitats. Presented an underwater fishing system using a minimal model. Images can be preprocessed with high-quality ND filters, which are different from regular ND filters. The ConvNet implementation was first trained on the large-scale ImageNet dataset. It was then refined and trained using sample images from the Fish4Knowledge dataset, achieving an accuracy of 85.05% [14]. Considered single-resolution images and developed a system to capture low-resolution images and convert them into high-resolution images. Machine learning engines extract features such as PCA-Net and Network-Intra-Network (NIN) [15]. The selected data is FishCLEF2015, and Katie's feature extraction consists of two Convolutional PCA layers of a deep learning network. The synchronization layer

contains a binary hash and a block histogram. Spatial pyramid pool follows this and extracts pose-invariant information. Classification is performed using linear SVM. The system achieved 98.64% accuracy on the Fish4Knowledge dataset [17]. Salman et al showed that CNNs use a combination of models to detect type- and dissimilarity-related features [14]. Muhammad et al presented an SVM architecture and feature extraction method for fish classification. SIFT and SURF are methods used to extract image features and achieve good results [21]. The above method uses machine learning techniques to identify fish species [8]. Conversely, other machines use traditional methods to accomplish the same task [9]. Provides an overview of the different methods used for fish classification [19]. In the literature, fish species classification is based on deep neural networks [22]. A multilayer neural network was used as a classifier to process features obtained from fish images [20]. This feature is based on parameters such as the size and shape of the fish [20]. The current method can be viewed as a special extractor of the input image and a discriminator of the extracted objects and proposes a method for the entire process of detecting and locating fish in images, which is defined as a special method using state-of-the-art technology [22]. Object detector type in Deep R-CNN [22]. In the literature, detection and classification of underwater fish are performed in unobstructed areas [11]. For example, we have a camera system installed on site [12]. The main task of species classification is mainly based on available data on large fish, including tropical fish [26]. The most important thing is not to focus on it [26]. Select external mode. The challenges of automated fish identification and classification systems are scalability and reliability [14]. Existing image processing methods sometimes fail to achieve their original purpose [14]. Therefore, it will be difficult to use it on the new platform in noisy and heavy situations [14]. Big data models attempt to eliminate the impact of data changes [14]. However, they must be large annotated datasets and require a lot of computing power [14]. The PC process, including the classification process, is shown in Table 1 [14].

Paper organization

- **Classification and Overview**
- **Pre-processing, training And dataset-related techniques.**
 - Pre-processing the input image
 - Applying data augmentation
 - Applying transfer learning
 - Comparing different training optimizers and datasets
 - Reducing annotation efforts
- **CNN design and optimization techniques.**
 - Designing collaborative or ensemble models
 - Combining the features of different layers and scales
 - Combining multiple types of features
 - Techniques for fine-grained classification
 - Selecting the best CNN
- **Conclusion remarks.**

This table lists the algorithms for three steps: preprocessing, extraction, and classification. The first step involves resizing the fish image, editing anchor points, and cropping the image.



SIFT
and

SURF distance measurement methods are often used for feature extraction. Many video extraction methods jointly extract size, texture, and color characteristics based on a combination of images. Commonly used fish classification algorithms include SVM, Bayesian classifier, and CNN [14]. Among these methods, SVM produces the best results for most data [14]. Automatic classification problem of underwater images. Automatic classification of underwater images is important, but it also presents many challenges [14]. When light is emitted, its intensity decreases due to energy loss [13]. These defects can reduce visibility and cause instability, especially in deep waters [13]. Additionally, compared to bodies of water such as lakes and ponds, the ocean has ocean currents that can change the brightness [25]. These changes, along with debris and debris removal, introduce noise into underwater images, especially ocean images [14]. Additionally, these images have low contrast and weak edges and details [14]. Colors are also distorted with distance due to uneven spectral distribution [14]. Reducing some of the limitations requires complex and expensive cameras [14]. The underwater images are endless, and the backgrounds are complex [14]. For example, the color of a sea cucumber is almost identical to the color of the surrounding sediment and algae [14]. This feature will help you stand out from the background [14]. Fish swim freely and quickly in three-dimensional space [14]. They love to hide behind other fish, coral, or sand [14]. Therefore, it is difficult to determine pose/orientation/size and perform semantic segmentation [14]. When there is a big change in the combination of animals and plants, the background also changes [14]. The front also appears to be altered due to the transparency, color, and solubility of the material [14]. Visualizing plankton distribution is difficult due to the large number of phylogenetic species present in plankton images and the small size of the organisms that produce them [17]. During underwater navigation, the diver's gestures must be recognized in different locations/situations and at different distances and angles [28]. Gestures are also difficult to distinguish from normal driving [28]. Similarly, coral reef datasets are rich in non-coral elements and limited in coral elements [32]. Additionally, although differences between different coral reef groups are unclear, there are significant differences within groups [32]. This reinforces the need for detailed classifications, which receive less attention than coarse classifications [32]. Underwater imaging systems have lower resolution than terrestrial imaging systems [33]. As such, it has special properties that can impact both people and technology [40].

IV. PROPOSED METHODOLOGY

This article explains what it takes to distinguish between several kinds of fish angles. The initial act is removing noise from the dataset. The use of pre-processed photos helps to mitigate the effects of non-fish items, foot water, and soil. Convolutional neural networks (CNN) subsequent use in deep learning methods for fish categorization. For CNN recognition and training to be as effective as possible, input images must include more features than the training samples. This succinct description sheds light on the procedures needed to correctly categorize fish angles.



Figure 1: Otsu's Binarization, dilation, and erosion pre-processing.

Preprocessing entails a number of crucial stages. First, Otsu thresholding is used to create a grayscale histogram in order to remove noise. The fish-centered face is clearly visible in pixels that are above the threshold, where some are seen as white and others as black. Morphological processes like erosion and dilatation then take place, changing the contours of the image. The image is then overlaid with a key overlay that indicates pixels based on a wrapped-up image, so decreasing the thickness of the foreground item. When at least one seed pixel is 1, expansion guarantees that a pixel in the final image is also 1, which is helpful for fragmented images. This technique efficiently gets images ready for further classification, improving accuracy by lowering noise and focusing on certain features in the data.

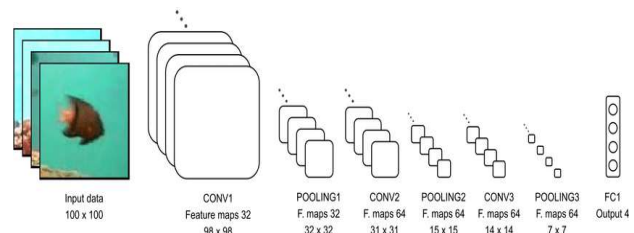


Figure 2: CNN's architecture

After denoising, fish are classified using a convolutional neural network (CNN) in the next step. The inputs to the camera are 100x100x3 raw RGB imagery overlaid with the 100x100x1 output from the previous step. These 100x100x4 inputs link to the hidden functions and receiver layer. The accumulation of gradients establishes weight, which makes object recognition possible regardless of position, and dense distribution prevents overfitting, which helps the CNN handle the issue successfully. Non-linear subsampling is carried out via the max-pooling layer, as shown in Figure (2). The layer's output is obtained using convolution, which combines indirect temporal and learning functions. For example, a 5x5x4 mask applied to a set of 100x100x4 requirements produces, when 32 5x5x4 filters are applied, a matrix with dimensions of 96x96x32. This method improves image analysis accuracy and efficiency by streamlining the classification of fish.

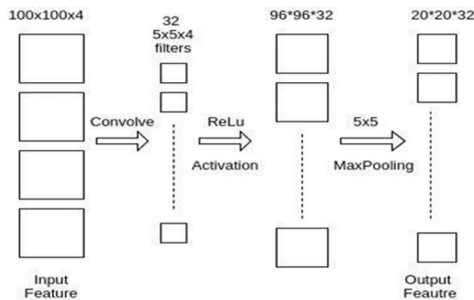


Figure 3: Max-pooling and 32 filters have been utilized to process the input feature.

The subsampling layer is the next element, which comes after the convolution layer and is located on the same plane. Its job is to reduce the size of the feature map by maximizing pooling and segmenting the image into 5x5 squares. This method helps with feature reduction while maintaining important data by keeping relative information between objects rather than their relationships. By reducing the size of the feature representation, subsampling improves computational efficiency and helps to speed up processing and lower memory usage. As a result, this layer is essential for network speed optimization since it makes feature extraction and analysis easier.

The request function implementation using maximum pooling and 32 filters is shown in Figure 3. Twice in this process, an iteration occurs: once with 64 filters and again with 32 filters. The final output then links to a fully connected layer, a distribution layer that is 80% linked, and yet another fully connected layer. Extending the image into appropriate groups through additional layers improves the accuracy of feature extraction and classification. This network structure enhances the extraction of pertinent features and allows for effective information transfer across the levels. As a result, this architecture enhances the network's capacity to recognize complex patterns and generate precise predictions.

To reduce the error between the network's demand and output, an algorithm is learned by the network. This technique contrasts several functions applied to various CNN techniques.

After full integration, the unemployment rate that is employed is:

In mathematics, cross entropy, represented as $H(y) = -\sum_i (y_i \log(y_i) + (1-y_i) \log(1-y_i))$, can be defined as follows: Y_i (negative) model log-like data entering.

V. CONCLUSION

The proposed angular species classification approach shows a remarkable accuracy of 96.29%, outperforming other current procedures for this application. This high accuracy combined with a compute-free time of 0.00183 seconds per loop makes this strategy suitable for real-time applications. However, problems remain while waiting to achieve 100% accuracy. It is mainly caused by noise from foundations and other water bodies which affects the classification of certain images. The goal is to improve the computation by implementing an image enhancement strategy to remove detected highlight defects in the image. Despite the challenges associated with underwater situations, especially uncontrolled environments, this paper highlights the central role of computer-assisted angle classification for marine researchers. Essentially, this paper highlights a gap in writing about latent image updating procedures moving from traditional machine learning to deep learning approaches. Analyst for a long time. While recognizing the model's validity for real-time applications, the study authors also recognized the need for ongoing change. In the future, it is proposed to combine image updating methods and information augmentation procedures using deep generative systems. Additionally, this study highlights the importance of large-scale, standardized datasets that expand historical angle species to include other marine species. The paper also advocates a new CNN scheme geared towards latent image classification and calls for considering logical spurious ideas to improve the interpretability of the show. The evolving nature of deep learning in latent research is spurring joint efforts from both academia and industry to transform this field into an entirely new field. In conclusion, this idea envisions solving these problems and improving the quality and security of deep networks, highlighting the continuous improvement of deep learning in the field of underwater image exploration.

VI. REFERENCES

- [1] S. Chen, B. Mulgrew, and P. M. Grant, "A clustering technique for digital communications channel equalization using radial basis function networks," *IEEE Trans. on Neural Networks*, vol. 4, pp. 570-578, July 1993.
- [2] J. U. Duncombe, "Infrared navigation—Part I: An assessment of feasibility," *IEEE Trans. Electron Devices*, vol. ED-11, pp. 34-39, Jan. 1959.
- [3] C. Y. Lin, M. Wu, J. A. Bloom, I. J. Cox, and M. Miller, "Rotation, scale, and translation resilient public watermarking for images," *IEEE Trans. Image Process.*, vol. 10, no. 5, pp. 767-782, May 2001.
- [4] White, D. J., Svelling, C., Strachan, N. J. C., *Automated measurement of species and length of fish by computer vision*, Elsevier, 2006.
- [5] F. Storbeck and B. Daan, *Fish species recognition using computer vision and a neural network*, Elsevier, 2000.
- [6] Nery, M. S., Machado, A. M., Campos, M. F. M., Padua, F. L. C., Carceroni, R., Queiroz-Neto, J. P., *Determining the appropriate feature set for fish classification tasks*, *Proceeding of the XVIII Brazilian Symposium on Computer Graphics and Image Processing*, 2005.
- [7] M. C. Chuang, J. N. Hwang, C. Rose, *Aggregated Segmentation of Fish from Conveyor Belt Videos*, *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1807-1811, 2013.

- epipelagics
in the
northern Gulf of Mexico using airborne lidar. *Bull. Mar.Sci.*2017, 93,591–609.
- [8] N. Castignolles, M. Cateon, M. Larinier, Identification and counting of live fish by image analysis, *SPIE. Vol 2182, Image and Video Processing II*, 1994.
 - [9] K.A. Mutasem, B.O. Khairuddin, N. Shahrulazman and A. Ibrahim (2010). "Fish Recognition Based on Robust Features Extraction from Size and Shape Measurements Using Neural Network". *Journal Computer Science*, Volume 6, Issue 10
 - [10] C. Spampinato, D. Giordano, R. D. Salvo, Y-H C-Burger, R.B. Fisher, G. Nadarajan, Automatic Fish Classification for Underwater Species Behaviour Understanding
 - [11] Y. Nagashima, Takakazu Ishimatsu, A Morphological Approach to Fish Discrimination, *IAPR Workshop on Machine Vision APPLICATIONS*, Nov. 17-19,1998
 - [12] B. Benson, J. Cho, D. Goshorn, R. Kastner, Field Programmable Gate Array (FPGA) Based Fish Detection Using Haar Classifiers, *American Academy of Underwater Sciences*, March 1, 2009.
 - [13] A. Rova, G. Mori, L. M. Dill, One fish, two fish, butterfly, trumpeter: Recognising fish in underwater videos, *IAPR Conference on Machine Vision Applications*.
 - [14] Moniruzzaman, M., Islam, S., Bennamoun, M., and Lavery, P., "Deep learning on underwater marine object detection: A survey," in *International Conference on Advanced Concepts for Intelligent Vision Systems*, 2017, pp. 150–160.
 - [15] L. Meng, T. Hirayama, and S. Oyanagi, "Underwater-drone with panoramic camera for automatic fish recognition based on deep learning," *IEEE Access*, vol. 6, pp. 17 880–17 886, 2018
 - [16] X. Li, Y. Tang, and T. Gao, "Deep but lightweight neural networks for fish detection," in *OCEANS 2017 - Aberdeen*, June 2017, pp. 1–5.
 - [17] A. Joly, H. Goeau, H. Glotin, C. Spampinato, P. Bonnet, W. P. Vellinga, " R. Planque, A. Rauber, S. Palazzo, B. Fisher, and H. M' uller, "Lifeclaf " 2015: Multimedia life species identification challenges," in *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, J. Mothe, J. Savoy, J. Kamps, K. Pinel-Sauvagnat, G. Jones, E. San Juan, L. Capellato, and N. Ferro, Eds. Cham: Springer International Publishing, 2015, pp. 462–483.
 - [18] E. Hossain, S. M. S. Alam, A. A. Ali and M. A. Amin, "Fish activity tracking and species identification in underwater video," 2016 5th International Conference on Informatics, Electronics and Vision (ICIEV), 2016, pp. 62 66, doi: 10.1109/ICIEV.2016.7760189.
 - [19] Kutlu, Y., Iscimen, B., Turan, C., 2017. Multi-stage fish classification system using morphometry. *Fresenius Environ. Bull.* 26, 1911–1917.
 - [20] Ali-Gombe A., Elyan E., Jayne C. (2017) Fish Classification in Context of Noisy Images. In: Boracchi G., Iliadis L., Jayne C., Likas A. (eds) *Engineering Applications of Neural Networks. EANN 2017. Communications in Computer and Information Science*, vol 744. Springer, Cham.
 - [21] G. Chen, P. Sun and Y. Shang, "Automatic Fish Classification System Using Deep Learning," 2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI), 2017, pp. 24-29, doi: 10.1109/ICTAI.2017.00016.
 - [22] Kratzert, F., Mader, H. Fish species classification in underwater video monitoring using Convolutional Neural Networks, *OpenKratzert, Frederik, and Helmut Mader. Fish Species Classification in Underwater Video Monitoring Using Convolutional Neural Networks. EarthArXiv*, vol. 15, 2018.
 - [23] Chang,C.;Fang,W.;Jao,R.C.;Shyu,C.;Liao,I.C.Development of an intelligent feeding controller for indoor intensive culturing of eel.*Aquac.Eng.*2005,32,343–353.
 - [24] Cabreira,A.G.; Tripode,M.; Madirolas, A.Artificial neural networks for fish-species identification. *ICESJ. Mar. Sci.* 2009, 66,1119–1129.
 - [25] Churnside,J.H.; Wells,R.; Boswell,K.M.; Quinlan,J.A.; Marchbanks,R.D.; McCarty,B.J.; Sutton, T.T.Surveying the distribution and abundance of flying fishes and other epipelagics in the northern Gulf of Mexico using airborne lidar. *Bull. Mar.Sci.*2017, 93,591–609.
 - [26] Jalali,M.A.; Ierodiaconou,D.; Monk,J.; Gorfine,H.; Rattray, A.Predictive mapping of abalone fishing grounds using remotely sensed LiDAR and commercial catch data. *Fish. Res.* 2015,169,26–36.
 - [27] R. K. Cowen and C. M. Guigand, "In situ ichthyoplankton imaging system (ISIIS): system design and preliminary results," *Limnology and Oceanography: Methods*, vol. 6, no. 2, pp. 126–132, 2008.
 - [28] "PlanktonSet dataset," <https://github.com/Planktos/PlanktonSet-1.0>, 2015.
 - [29] "Fish4Knowledge dataset," <http://groups.inf.ed.ac.uk/f4k/>, 2015.
 - [30] "Fish-gres dataset," <https://data.mendeley.com/datasets/76cr3wfhff/1>, 2021.
 - [31] O. Beijbom, P. J. Edmunds, D. I. Kline, B. G. Mitchell, and D. Krieg man, "Automated annotation of coral reef survey images," in *CVPR*, 2012, pp. 1170–1177.
 - [32] A. Gomez Chavez, A. Ranieri, D. Chiarella, E. Zereik, A. Babi' c, and A. Birk, "CADDY underwater stereo-vision dataset for human–robot interaction (HRI) in the context of diver activities," *Journal of Marine Science and Engineering*, vol. 7, no. 1, p. 16, 2019.
 - [33] A. G. Howard, "Some improvements on deep convolutional neural network based image classification," *arXiv:1312.5402*, 2013.
 - [34] P. Pudil, J. Novovi' cov'a, and J. Kittler, "Floating search methods in feature selection," *Pattern recognition letters*, vol. 15, no. 11, pp. 1119 1125, 1994.
 - [35] A. Lumini and L. Nanni, "Ocean ecosystems plankton classification," in *Recent Advances in Computer Vision*. Springer, 2019, pp. 261–280.
 - [36] H. Zheng, R. Wang, Z. Yu, N. Wang, Z. Gu, and B. Zheng, "Automatic plankton image classification combining multiple view features via multiple kernel learning," *BMC bioinformatics*, vol. 18, no. 16, p. 570, 2017.
 - [37] L. Liu, C. Shen, and A. van den Hengel, "The treasure beneath convolutional layers: Cross-convolutional-layer pooling for image clas sification," in *CVPR*, 2015, pp. 4749–4757.
 - [38] P. Li, J. Xie, Q. Wang, and W. Zuo, "Is second-order information helpful for large-scale visual recognition?" in *IEEE ICCV*, 2017, pp. 2070–2078.
 - [39] S. Dieleman, J. Fauw, and K. Kavukcuoglu, "Exploiting cyclic symmetry in convolutional neural networks," 02 2016.
 - [40] X. Sun, J. Shi, L. Liu, J. Dong, C. Plant, X. Wang, and H. Zhou, "Transferring deep knowledge for object recognition in low-quality underwater videos," *Neurocomputing*, vol. 275, pp. 897–908, 2018.