Learning Visual Representation of Underwater Acoustic Imagery Using Transformer-Based Style Transfer Method

Xiaoteng Zhou
School of Ocean Engineering,
Harbin Institute of Technology
zhouxiaoteng525@gmail.com

Changli Yu∗
School of Ocean Engineering,
Harbin Institute of Technology
yuchangli@hitwh.edu.cn

Shihao Yuan
School of Computer Science and Technology,
Harbin Institute of Technology
yuanshihao1014@gmail.com

Xin Yuan
School of Ocean Engineering,
Harbin Institute of Technology
xin.yuan@upm.es

Hangchi Yu
School of Ocean Engineering,
Harbin Institute of Technology
yuhangchi@163.com

Citong Luo
School of Ocean Engineering,
Harbin Institute of Technology
luocitong@gmail.com

ABSTRACT
Underwater automatic target recognition (UATR) has been a challenging research topic in ocean engineering. Although deep learning brings opportunities for target recognition on land and in the air, underwater target recognition techniques based on deep learning have lagged due to sensor performance and the size of trainable data. This letter proposed a framework for learning the visual representation of underwater acoustic imageries, which takes a transformer-based style transfer model as the main body. It could replace the low-level texture features of optical images with the visual features of underwater acoustic imageries while preserving their raw high-level semantic content. The proposed framework could fully use the rich optical image dataset to generate a pseudo-acoustic image dataset and use it as the initial sample to train the underwater acoustic target recognition model. The experiments select the dual-frequency identification sonar (DIDSON) and adaptive resolution imaging sonar (ARIS) as the underwater acoustic data source and also take fish, the most common marine creature, as the research subject. Experimental results show that the proposed method could generate high-quality and high-fidelity pseudo-acoustic samples, achieve the purpose of aquatic data augmentation and provide support for the underwater acoustic-optical images domain transfer research.

CCS CONCEPTS
• Computing methodologies; • Computer vision tasks;

KEYWORDS
Ocean engineering, underwater acoustic imagery, sonar data augmentation, style transfer, deep learning

∗Corresponding author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ICDIP 2023, May 19–22, 2023, Nanjing, China
© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-9825-7/23/05...$15.00
https://doi.org/10.1145/3604078.3604176

1 INTRODUCTION
As the human exploration of the ocean continues, more and more sensors are being developed to acquire rich underwater data. Due to the absorption and refraction of light by the seawater medium, optical cameras are limited in their range of action. The cameras usually need to be equipped with high power-consuming external light source devices, which leads to the limited application space underwater. The limitations of optical sensors are significantly magnified in deep water environments. In contrast, sonar sensors have become the primary sensors in underwater detection tasks due to their long detection distance and freedom from turbid zone interference [1]. With the development of the electronics industry, imaging sonars with high resolution have been successfully developed, such as dual-frequency identification sonar (DIDSON) and adaptive resolution imaging sonar (ARIS), which belong to a class of forward-looking sonar (FLS). DIDSON and ARIS are also called acoustic cameras because their imaging effects are close to those of optical cameras, which are better than mechanical scanning sonars. In addition, the installation and operation of DIDSON and ARIS sensors are more flexible than side scan sonars (SSSs) and synthetic aperture sonars (SASs), which could significantly reduce imaging errors caused by carrier platform motion. High-performance imaging sonar improves the benefits of underwater acoustic vision applications, accelerates the development of acoustic image-based ocean engineering applications, and lays the foundation for the introduction of deep learning tools. Currently, underwater acoustic images acquired by DIDSON and ARIS are widely used in fisheries, underwater archaeology, structure maintenance and ecological environment monitoring [2–4].

Traditional underwater automatic target recognition (UATR) tasks usually require three parts: pre-processing, feature extraction, and feature classification; however, most of the methods used in these three parts are developed for optical images and are not
compelling enough when applied to underwater acoustic scenes. These traditional methods also rely heavily on researchers’ priori knowledge in parameter tuning. A good solution usually requires a lot of trial and error, which is not friendly enough for newcomers in the underwater vision field. With the rapid development of deep learning in computer vision, more and more target recognition tasks are performed using Convolutional Neural Networks (CNNs), and the recognition results have been widely validated \cite{5, 6}. In essence, these learning-based recognition algorithms adopt data-driven approaches. The models learn features and their underlying laws from a large number of labeled training samples and use them for classification and recognition, with the whole process approaching an end-to-end mode without complex manual design. Meanwhile, the deep learning-based approaches require less expert analysis and fine-tuning. The knowledge and expertise required to process underwater acoustic images manually are replaced by the knowledge and expertise required to iterate using deep learning architectures \cite{7}, a process that has changed as shown in Figure 1, which is in line with the trend towards intelligent underwater applications.

However, while deep learning brings opportunities to underwater acoustic target recognition, it also poses significant challenges. CNN-based approaches require a large amount of labeled training data upfront, and the features obtained from deep neural network learning are based on a specific training dataset. In short, if there is a problem constructing the datasets used for training, the network will become less effective in processing images outside the training datasets. However, due to the complex and unknown characteristics of the underwater environment, it is expensive and time-consuming to collect large amounts of underwater acoustic data.

In computer vision, some researchers have suggested the concept of transfer learning \cite{8}, which entails fine-tuning a model trained on high-volume non-target data with low-volume target samples to get superior prediction results finally. However, attempts at this idea tend to focus on the same domain, that is, the training image and the target image have less heterogeneity and common features, such as two optical imaging domains or two remote sensing imaging domains. Due to the different imaging methods used by optical cameras and imaging sonar in underwater situations, the final images obtained have a substantial domain offset problem \cite{9}, which makes transfer learning face a significant challenge in underwater acoustic target recognition tasks. Moreover, transfer learning requires a certain amount of initial target data, but it cannot be fully supported because most deepwater environments to be explored are usually harsh and complex, and most underwater targets are zero samples or extremely scarce. For this extreme data scarcity, some researchers have proposed synthesizing pseudo-samples, essentially a kind of data augmentation. Style transfer is a deep augmentation technique instead of data augmentation techniques in the conventional sense, such as rotation, cropping, or adding noise. It is based on the fundamental idea of keeping the high-level semantic information, also known as content, in the image and then replacing low-level texture features, also known as styles \cite{10}. Style transfer-based training data augmentation approaches are very flexible, unlike image simulation software-based and GAN network-based methods \cite{11–13}, as these methods often do not require experience from pre-preparation. The pseudo-acoustic sample generation methods based on style transfer are more suitable for scenarios where the initial samples are unknown, such as identifying unknown fish in uncharted waters. When a specific size of training samples is generated, they could be used to train recognition and classification models for underwater acoustic images, which will be more in-depth and flexible than the transfer learning-based training model.

Using style transfer approaches to generate underwater pseudo-acoustic image datasets and thus complete the initial training sample accumulation is effective \cite{14, 15}. However, their study was conducted for underwater SSS and multibeam sonar images. There is a restricted application space for the pseudo-acoustic datasets after style transfer since the images detected by these two sonar sensors are still visually distinct from the optical ones. Unlike them, this paper introduces a database of underwater images detected by the DIDSON and ARIS devices, which are visually very close to optical images, and it could be assumed that there are only domain
Learning Visual Representation of Underwater Acoustic Imagery Using Transformer-Based Style Transfer Method

ICDIP 2023, May 19–22, 2023, Nanjing, China

Figure 2: The overall research framework proposed in this paper.

Since most of the current automatic sonar image recognition tasks are implemented on private datasets, few publicly available models are directly used for practical underwater applications, which is one of the motivations for writing this paper. In this research, the proposed framework is validated using underwater acoustic images captured from random, natural, and unknown marine environments. The specific contributions are summarized below.

- A transformer-based style transfer approach is proposed to be introduced to underwater acoustic images and is optimally adapted.
- Generating pseudo-acoustic image datasets using a style transfer-based pipeline, then systematically evaluating and analyzing these datasets.
- The research is based on high-resolution acoustic data for fish identification tasks in unknown waters.

The remainder of this article is organized as follows. Section II describes the details of the method. Section III illustrates the experiment and validation methods. Section IV analyzes and evaluates the experimental results. Section V discusses the shortcomings of the experiments and gives some outlook. Conclusions are given in section VI.

2 METHODOLOGY

2.1 Style transfer model

In order to improve the practical application value of this research in marine engineering, the design of the method in this paper is not limited to a specific link. However, it proposed a complete research framework, aiming to directly connect the resource-rich optical images with the sample-poor acoustic images through this framework, where the optical images provide high-level semantic
information to exploit their imaging advantages effectively. In contrast, acoustic images only provide stylized textures to avoid their imaging limitations cleverly. Finally, the generated pseudo-sample dataset is used to train downstream task models, such as underwater target recognition and segmentation models. In the traditional transfer methods, it is frequently necessary to continuously deepen the number of network layers to improve the transfer effect. However, this will result in a reduction in the image’s resolution and the loss of feature details, which will produce an incomplete and inaccurate representation of the final transferred image content and prevent the style texture from being adequately embedded. For this problem, the approach of this paper mainly applies the latest research in the field of style transfer using transformer [18] to the field of sonar image generation and compares the gap with other methods. The model architecture developed in this paper has a better global feature transfer effect and more robust feature representation capability, which can minimize detail loss during the image transfer process while ensuring the structural integrity and higher quality of the generated pseudo-acoustic images. The basic idea of a transformer-based style transfer network adopts the Vision Transformer (ViT) approach, which is more widely used in computer vision, to improve the performances of the transformer on the style transfer task by proposing the spatially invariant positional encoding method content-aware positional encoding (CAPE), which avoids the sensitivity of transformer to scale changes while ensuring that transformer captures the long-range information of the image.

The transformer used in this model is originally a robust structure in natural language processing. Its basic structure consists of an encoder and decoder, where the encoder converts the sentence and position encoding composed of the token into intermediate encoding through the attention mechanism. Then, the decoder converts the intermediate encoding into output results; eventually, the sequence-to-sequence generation task is achieved. Both encoder and decoder have a self-attention mechanism to extract semantic information, and then the decoder’s attention mechanism to translate. The transformer can also perform the image-to-image generation task by decomposing the image into several patches and then encoding the position. By introducing CAPE, the structural information of the content image is preserved. In contrast, the traditional position coding of the style image loses the structural information so that only the style texture information is preserved. Then the intermediate coding of the two parts is mixed and decoded by the transformer’s decoder to form the image after style transfer.

This paper applies the transformer-based style transfer method to pseudo-sonar image generation. Since the style transfer algorithm could only transfer style and texture information, this work enhances the optical images before the transfer step to improve the final effects. The leading optimization considers that the sonar image is a single-channel colorized pseudo-color image, so the original optical image is grayed out to simulate a single-channel acoustic image. Meanwhile, considering that the background of the sonar image is usually featureless seawater, while our optical image is an especially captured fish image, we performed foreground extraction for the optical image to simulate the effect of an actual sonar image without a background. Here, the foreground extraction is performed using the image matting method, and a manual assistance method is used to make the effect more accurate. In addition, the necessary image processing, such as Gaussian smoothing, is also used appropriately to make the result of optical image transfer closer to the actual sonar image. The overall image processing pipeline is shown in Figure 3.

3 EXPERIMENTAL SETTINGS
This subsection describes the details of the experimental setup, the source of the training dataset, and the strategy for validating the experiments. The overall architecture is shown in Figure 4.

3.1 Content image dataset
In the stage of training the style transfer model, this manuscript introduces the marine fish dataset provided in the literature [19], which is rich in variety and contains variations in scale and orientation. It is initially used to train the image classifier and segmentation models. This research uses this marine fish dataset (as shown in Figure 5) to train a style transfer model, mainly considering that the visual presentation of these samples is very close to the bird’s eye view style presentation of sonar, which facilitates the rendering of optical images later using the transfer model. The model treats the marine fish dataset as image content for input, and the number of training samples collected is 9000, and some of the sample images are displayed as follows.

3.2 Style image dataset
The style image dataset input to the transfer model is derived from the image gallery of SOUND METRICS [4]. To ensure the model’s generalization ability, this manuscript does not classify the acoustic imaging styles because various ocean backgrounds will produce different acoustic image styles depending on the specific application scenario. This research performed frame decomposition on the images in GIF format in the dataset and finally collects 3952 training samples, some of which represent image styles as shown in Figure 6.

3.3 Verification methods
After completing the training of the style transfer model, this manuscript analyzes the transfer effects from qualitative and quantitative perspectives, in which DBCNN [20], NIQE [21], and BRISQUE [22] are introduced to evaluate the own quality of the generated pseudo-acoustic samples quantitatively. Cosine similarity [23] and Perceptual Hash [24] are used to quantitatively evaluate the similarity between the generated pseudo-acoustic samples and the authentic acoustic images. In addition, a comprehensive comparison between the method in this paper and the style transfer methods used in [10], [14], and [15] is also performed, and the comparison results are presented in schematic and tabular form.

3.4 Testing environment
All experiments were performed on a 64-bit PC with a Windows 10 operating system equipped with an Intel (R) Xeon (R) Gold 6226R 2.90 GHz processor, 128 GB of physical memory, and an NVIDIA Tesla T4 graphics card. Python 3.8 was used to compile the program.
Figure 3: Schematic diagram of the overall image processing pipeline.

Figure 4: The flowchart of the overall experimental design.

Figure 5: Examples of marine fish samples.
Figure 6: Examples of underwater acoustic imagery data.

Figure 7: The transferred effects diagram corresponding to different methods.

Table 1: Quality assessment results of pseudo-acoustic sample A generated by different transfer methods

| Metrics          | Gatys et al. [10] | Lee et al. [14] | Li et al. [15] | Ours  |
|------------------|-------------------|-----------------|---------------|-------|
| DBCNN (higher - better) | 29.3614           | 24.8291         | 14.3595       | 25.0076 |
| NIQE (lower - better)      | 10.7025           | 10.3499         | 7.3146        | 7.1349  |
| BRISQUE (lower - better)   | 91.2619           | 75.1481         | 32.5592       | 62.7893 |

4 EXPERIMENTAL RESULTS AND EVALUATION

4.1 Comparison with other style transfer methods

This subsection first compares the rendering performance with other typical style transfer methods. Considering the limitation of manuscript length, three samples of pseudo-acoustic images are randomly selected for demonstration in this part. The visual effect comparison of style transfer is shown in Figure 7, and the detailed quantification results are shown in Table 1, Table 2, and Table 3.

Figure 7 shows that the visual features of the pseudo-acoustic images generated by the transfer model are more apparent and more complete in structure, and the rendering effect is more in-depth. In contrast, other transfer methods will destroy the image details, and the rendering effects are unsatisfactory. From the statistical
Table 2: Quality assessment results of pseudo-acoustic sample B generated by different transfer methods

| Metrics Methods | DBCNN(higher - better) | NIQE(lower - better) | BRISQUE(lower - better) |
|-----------------|-------------------------|----------------------|-------------------------|
| Gatys et al. [10] | 28.7913                 | 10.5685              | 90.9425                 |
| Lee et al. [14]  | 21.8231                 | 9.8479               | 68.2902                 |
| Li et al. [15]   | 17.4809                 | 6.9175               | 40.8156                 |
| Ours             | 22.9797                 | 8.1695               | 62.7892                 |

Table 3: Quality assessment results of pseudo-acoustic sample C generated by different transfer methods

| Metrics Methods | DBCNN(higher - better) | NIQE(lower - better) | BRISQUE(lower - better) |
|-----------------|-------------------------|----------------------|-------------------------|
| Gatys et al. [10] | 25.4517                 | 10.2652              | 83.1222                 |
| Lee et al. [14]  | 20.8625                 | 7.6013               | 48.9831                 |
| Li et al. [15]   | 18.5751                 | 7.2107               | 45.1564                 |
| Ours             | 22.0630                 | 8.1983               | 58.2721                 |

Figure 8: The effects of style transfer in various acoustic scenes.

In the results of above table statistics, the method in this paper is not always optimal in some cases, which may be related to the excessive removal of image noise, because noise is a property of underwater acoustic images. Besides, the evaluation method is designed for optical images, and there are certain deviations in the evaluation results. Designing a dedicated quality evaluation method for underwater acoustic images will be one of the key extensions of this research.

4.2 Comparison of transfer effects in various acoustic scenes

This subsection demonstrated the rendering effects of our method in various underwater acoustic scenes. As seen in Figure 8, the method we proposed could be well adapted to complex acoustic scenes, and the style texture features provided by these scenes also could be well learned and embedded. In practical underwater engineering applications, the operator could choose the transfer mode according to the acoustic environment of the specific underwater tasks and finally achieve the purpose of high-precision image dataset generation.
4.3 Comparison with real underwater acoustic images

In this subsection, the style transfer model proposed is used to transfer practical fish imageries. The final generated pseudo-acoustic images are evaluated for similarity with the natural acoustic camera-captured images [16] to verify the method’s effectiveness in this paper. The comparison schematic is shown in Figure 9, and the similarity evaluation results are shown in Table 4.

From the statistical results in Table 4, it could be seen that the average similarity evaluation results of the pseudo-acoustic images generated and the real acoustic images exceed 0.75. The experimental findings demonstrate that the approach suggested in this letter could produce pseudo-acoustic images that are highly comparable to natural underwater acoustic imaging. It is important to note that, as illustrated in Figure 9, the pseudo-acoustic images produced by our method still retain the noisy features, which are an essential aspect of underwater acoustic imaging. The preservation of noise details also verifies the rendering effect and generalization ability of the model from a deep aspect. In subsequent work, we intend to train a marine fish recognizer using the generated pseudo-acoustic pictures and further assess the effects of style transfer using the recognition’s accuracy and error.

5 DISCUSSIONS

The purpose of acoustic-optical image style transfer is to make the feature distributions of the target and source domains closer, reduce the domain bias due to the difference in imaging mechanisms, and then transfer the knowledge information effectively between the two domains. From the experimental results, it can be seen that realistic underwater pseudo-acoustic samples can be generated using the transformer-based style transfer method. From a qualitative perspective, compared with other CNN-based methods, the visual effect of the pseudo-acoustic images generated using this mode is closer to the underwater acoustic detection effect, the image resolution is higher, and the details are better preserved. From the quantitative point of view, the generated pseudo-acoustic images...
have higher quality and higher similarity between them and authentic acoustic images, which can be directly used to train target recognition models. However, in the overall experimental process, the method proposed in this paper still has some shortcomings: (1) The style transfer method could effectively retain image details and better affect the style transfer of acoustic images. However, when transferring the high-level semantic content of optical images, the transfer of feature details is too clear, resulting in the detail discrimination of the final generated pseudo-acoustic images exceeding the DIDSON and ARIS sensors’ discriminative ability. Later, we will add random noise to make the generated pseudo-acoustic images closer to the authentic captured underwater acoustic images. (2) The transformer-based style transfer method requires high computing resources. It does not have advantages in speed, which hardly applies to portable computing devices at present, so the lightweight model will be one of our following research goals. (3) Due to the limitation of current experimental conditions, the sonar image and optical image datasets selected in this paper are not strictly aligned in terms of semantic content. In the future, we will detect more acoustic-optical images of fish from natural ocean scenes and study the effects of unilateral domain transfer after aligning the semantic information of acoustic-optical images, which will open new horizons for registration and reconstruction tasks based on acoustic-optical images.

6 CONCLUSIONS
This paper presents a framework for generating underwater pseudo-acoustic image datasets using a transformer-based style transfer method and it is tested on a natural marine fish dataset. The experimental results show that the quality of the pseudo-acoustic images generated by this method is better than the existing acoustic image transfer methods, and the average similarity between the pseudo-acoustic images generated and the authentic marine acoustic images is close to 0.8. In the future, to validate the proposed framework in practical marine engineering applications, we will use this approach to generate a sufficiently large dataset of underwater pseudo-acoustic images, and then, we will use this dataset to train an automated marine fish recognition model. In addition, we will simplify the framework to adapt it to the hardware to enhance the autonomous underwater operation capabilities of autonomous underwater vehicles.

ACKNOWLEDGMENTS
This work was supported by the Chinese Shandong Provincial Key Research and Development Plan, under Grant No. 2021CXGC010702.

REFERENCES
[1] Blondel, Philippe, and Bramley J. Murton. 1997. Handbook of seafloor sonar imagery. Vol. 7. Chichester: Wiley.
[2] Hurtos, Natalia, David Ribas, Xavier Cufi, Ivan Pettitot, and Joaquim Salvi. 2015. Fourier-based registration for robust forward-looking sonar mosaicing in low-visibility underwater environments. Journal of Field Robotics 32(1) (2015), 123-151. https://doi.org/10.1002/rob.21516
[3] Negahdarpour and Shahriar. On 3-D motion estimation from feature tracks in 2-D FS sonar video. 2013. IEEE Transactions on Robotics 29:4 (2013), 1016-1030.
[4] Sound Metrics Dataset. Retrieved September 6, 2022 from http://www.soundmetrics.com.
[5] Teng, Bowen and Hongjian Zhao. 2020. Underwater target recognition methods based on the framework of deep learning: A survey. International Journal of Advanced Robotic Systems 17(6). https://doi.org/10.1772/jirass14067907307.
[6] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. 2017. Imagenet classification with deep convolutional neural networks. Communications of the ACM 60(6), 84-90. https://doi.org/10.1145/3065386.
[7] O’Mahony, Niall, Sean Campbell, Anderson Carvalho, Sunan Harapanahalli, Gustavo Velaeso Hernandez, Lenka Krapalkova, Daniel Bierdzan, and Joseph Walsh. 2019. Deep learning vs. traditional computer vision. In Science and information conference. Springer. Cham, 128-144. https://doi.org/10.1007/978-3-030-17959-9_10.
[8] Zhuang, Funzh, Zhiyuan Q., Keyu Duan, Dongbo Yu, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qiong He. 2020. A comprehensive survey on transfer learning. In Proceedings of the IEEE 109(1) 43-76. https://doi.org/10.1109/JPROC.2020.3004555.
[9] Quinonero-Candela, Joaquin, Masashi Sugiyama, Anton Schwaighofer, and Neil D. Lawrence. 2022. Dataset shift in machine learning. MIT Press, USA.
[10] Gatsys, Leon A., Alexander S. Ecker, and Matthias Bethge. 2016. Image style transfer using convolutional neural networks. In Proceedings of the IEEE Conference on computer vision and pattern recognition (CVPR). IEEE, Las Vegas, 2414-2423. https://doi.org/10.1109/cvpr.2016.265.
[11] Sung, Minsung, Jason Kim, Mungiuk Lee, Byeonjin Kim, Taesik Kim, Juwhan Kim, and Son-Choel Yu. 2020. Realistic sonar image simulation using deep learning for underwater object detection. International Journal of Control, Automation and Systems 18(3), 523-534. https://doi.org/10.1007/s12555-019-0691-3.
[12] Liu, Dingyu, Yusheng Wang, Yonghong J., Hiroshi Tsuchiya, Atsushi Yamashita, and Hajime Asama. 2021. CycleGAN-based realistic image dataset generation for forward-looking sonar. Advanced Robotics 35, 242-254. https://doi.org/10.1080/01691864.2021.1873845.
[13] Rixon Fuchs, Louise, Christr Larsson, and Andreas Gallstroem. 2019. Deep learning based technique for enhanced sonar imaging. In 5th Underwater Acoustics Conference & Exhibition (UACE), Hersonissos, Crete, Greece. Jun. 30 2019-Jul. 5 2019, 1021-1028.
[14] Lee, Sejin, Byungsae Park, and Ayoung Kim. 2019. Deep learning based object detection via style-transfered underwater sonar images. IFAC-PapersOnLine 52(21), 152-155. https://doi.org/10.1016/j.ifacol.2019.12.289.
[15] Li, Chuanlong, Xiufen Ye, Dongxian Cao, Jie Hou, and Haibo Yang. 2021. Zero-shot objects classification method of side scan sonar image based on synthesis of pseudo samples. Applied Acoustics 173 (2021): 107691. https://doi.org/10.1016/j.apacoust.2020.107691.
[16] Zhang, Y., K. Mizuno, A. Asada, Y. Fujimoto, and T. Shimada. 2016. New method of fish classification by using high-resolution acoustic video camera-ARIS and local invariant feature descriptor. In Proceedings of OCEANS. https://doi.org/10.1109/OCEANS.2016.7761115.
[17] Schneider, Stefan, and Alex Zhang. 2020. Counting fish and dolphins in sonar images using deep learning. ArXiv preprint. ArXiv:2007.12808. https://doi.org/10.48550/arXiv.2007.12808.
[18] Deng, Yingying, Fan Tang, Weiming Dong, Chongyang Ma, Xingjia Pan, Lei Wang, and Changsheng Xu. 2022. StyTr2: Image Style Transfer with Transformers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, New Orleans, LA, 11326-11336. https://doi.org/10.1109/CVPR52688.2022.01104.
[19] Ulucan, Oguzhan, Diclehan Karakaya, and Mehmet Turkan. 2020. A Large-Scale Dataset for Fish Segmentation and Classification. In 2020 Innovations in Intelligent Systems and Applications Conference. IEEE, Istanbul, 1-5. https://doi.org/10.1109/INISIS5.2020.9259867.
[20] Zhang, WeiXia, Kede Ma, Jia Yan, Daxiang Deng, and Zhou Wang. 2018. Blind image quality assessment using a deep bilinear convolutional neural network. IEEE Transactions on Circuits and Systems for Video Technology 30(1) 36-47. https://doi.org/10.1109/TCSVT.2019.2868671.
[21] Mittal, Anish, Rajiv Soundararajan, and Alan C. Bovik. 2012. Making a ‘completely blind’ image quality analyzer. IEEE Signal processing letters 20(3), 209-212. https://doi.org/10.1109/LSP.2012.227722.
[22] Mittal, Anish, Anush Krishna Moorthy, and Alan Conrad Bovik. 2012. No-reference image quality assessment in the spatial domain. IEEE Transactions on image processing 21(12), 4695-4708. https://doi.org/10.1109/TIP.2012.2240540.
[23] Liu, Chengjun. 2014. Discriminant analysis and similarity measure. Pattern Recognition 47(1), January, 2014, 359-367. https://doi.org/10.1016/j.patcog.2013.06.023.
[24] Shao-zhong, Wang, and Zhang Xin-peng. 2007. Recent development of perceptual image hashing. Journal of Shanghai University (English Edition) 11(4), 323-331.