Underwater Image Super-Resolution using Generative Adversarial Network-based Model

Alireza Aghelan, Modjtaba Rouhani
Computer Engineering Department, Ferdowsi University of Mashhad
Mashhad, Iran
aghelan.alireza@mail.um.ac.ir, rouhani@um.ac.ir

Abstract—Single image super-resolution (SISR) models are able to enhance the visual quality of underwater images and contribute to a better understanding of underwater environments. The integration of these models in Autonomous Underwater Vehicles (AUVs) can improve their performance in vision-based tasks. Real-ESRGAN is a powerful model that has shown remarkable performance among SISR models. In this paper, we optimize the Real-ESRGAN model for underwater image super-resolution. To optimize and evaluate the performance of the model, we use the USR-248 dataset. The proposed model generates images that demonstrate a higher level of visual quality than the outputs of the Real-ESRGAN model.

Keywords—Underwater images; Single image super-resolution; Deep learning; Generative adversarial network

I. INTRODUCTION

Autonomous Underwater Vehicles (AUVs) serve critical roles in various significant domains, such as the monitoring of marine species, collaboration between humans and robots, analyzing underwater environments, mapping the seabed, and more [1]. To successfully accomplish their objectives, which involve tasks like tracking and comprehending scenes, these vehicles heavily rely on high-quality images. By using precise visual information, AUVs can make informed decisions, navigate complex underwater environments, and contribute to the success of diverse underwater missions.

The quality of underwater images is often negatively affected due to various factors, such as absorption, scattering, and limited visibility. These factors lead to the loss of significant details and the production of images that appear blurred. Despite the use of advanced cameras, this problem remains unresolved, presenting difficulties in this domain. The adoption of rapid and precise image super-resolution models can contribute to resolving this problem by enhancing the resolution and visual quality of underwater images. [1]

The scarcity of underwater datasets encompassing high-quality and high-resolution (HR) images poses a noteworthy challenge in the field of underwater image super-resolution. Obtaining such datasets is challenging due to the difficulties associated with capturing clear and detailed underwater scenes. However, in this work, we address this challenge by using the USR-248 dataset, which comprises a collection of good-quality underwater images. By utilizing this dataset, we can enhance the robustness and effectiveness of our proposed method in producing HR outputs with improved visual quality.

Another important challenge is the limited amount of data in most underwater datasets. These datasets contain only a few hundred images, which is often insufficient for models to learn the intricate patterns in underwater scenes. This limitation makes it difficult to train complex models from scratch. To overcome this challenge, we use the transfer learning technique. By adopting this approach, we leverage existing knowledge from pre-trained models to improve our models' performance.

In dynamic underwater environments, image processing speed can greatly impact an AUV's ability to perform tasks such as obstacle avoidance, navigation, and rapid decision-making. Therefore, it is important to note that the underwater image super-resolution models should have high computational efficiency and rapid inference time. This ensures that AUVs utilizing these models can deliver effective and dependable performance when processing underwater images, enabling them to effectively perform their tasks.

Traditional techniques for single image super-resolution (SISR), including interpolation-based and reconstruction-based methods, exhibit limited efficacy in recovering high-frequency information. They do not consider the complex non-linear mapping between low and high-resolution images. These methods usually produce blurry images that lack fine textures. Currently, deep learning-based methods for SISR have become prevalent and demonstrate superior performance compared to traditional approaches. These approaches have the ability to learn to reconstruct photo-realistic details from LR images rather than relying on mathematical models. Among deep learning-based SISR techniques, approaches based on Generative Adversarial Networks (GAN) [2] have exhibited substantial improvement, and many studies have been conducted on them.

GAN-based super-resolution models utilize generator and discriminator networks in an adversarial training process. The generator network aims to produce HR outputs indistinguishable from real images. Simultaneously, the discriminator distinguishes between the real images and the generated images, forcing the generator to produce higher-quality outputs. This adversarial training process enables the model to learn the underlying patterns in the data and generate high-resolution images with improved visual quality.

To optimize the Real-ESRGAN model for underwater image super-resolution, we adopt the transfer learning technique. By leveraging existing knowledge from pre-trained models, we can enhance the performance of our proposed method. This approach not only improves the quality of the generated images but also reduces the computational cost and training time.

In this paper, we present our approach for underwater image super-resolution using the Real-ESRGAN model and the USR-248 dataset. We evaluate the performance of our proposed model and compare it with other state-of-the-art SISR methods. The results demonstrate the effectiveness of our approach in enhancing the visual quality of underwater images, making them suitable for various underwater missions.
samples, providing valuable feedback to the generator. This adversarial competition drives the generator to synthesize realistic texture details, resulting in a significant improvement in both visual realism and overall quality of the HR images compared to previous approaches.

The Real-ESRGAN model [3] is extensively employed as a practical solution for general image super-resolution tasks. This model has shown excellent performance in recovering HR images from real-world low-resolution (LR) images. Furthermore, the computations in this model are performed at an acceptable speed. Inspired by these strengths, we were motivated to fine-tune the Real-ESRGAN model using an underwater dataset and leverage this model to improve the visual quality of underwater images.

II. RELATED WORK

This section provides an overview of several works in the field of underwater image super-resolution alongside a summary description of the Real-ESRGAN model.

The degradation of underwater images can be attributed to factors like scattering. Lu et al. [4] proposed a self-similarity-based technique for de-scattering and resolution enhancement of underwater images. Traditional techniques often result in the loss of high-frequency details during the de-scattering. To tackle this problem, the authors introduced a high turbidity underwater image super-resolution approach. The proposed model produces pleasant outputs with a reasonable noise level.

Islam et al. [1] developed SRDRM and SRDRM-GAN models to perform underwater image super-resolution, specifically designed for autonomous underwater robots. The SRDRM model is a generative model based on a deep residual network architecture. In the SRDRM-GAN model, the SRDRM model performed the role of the generator, and a Markovian PatchGAN-based model operated as the discriminator. In addition, the authors incorporated an objective function to guide the training process. The objective function assesses the perceptual quality of an image by taking into account its overall content, color, and texture information. They also presented the USR-248 dataset that we use in our work. Details about the USR-248 dataset can be found in Section 3.

AlphaSRGAN [5] is an efficient GAN-based underwater image super-resolution model that merges traditional image reconstruction techniques with deep learning methods. The authors incorporate pre-processing images before entering them into the generator network, resulting in improved performance and stability. In their work, the USR 248 dataset [1] is utilized for training and evaluation. This model demonstrates remarkable performance and is well-suited for real-time applications.

In Wang et al. [6] the authors proposed a lightweight multi-stage information distillation network to balance computational speed and model performance in underwater image super-resolution tasks. The researchers introduced a recursive residual feature distillation module to extract features using a minimal parameter count. Additionally, they employed a channel interaction and distillation module to extract valuable information without requiring additional parameters. These modules prove highly beneficial in reducing the consumption of computing resources.

Based on the successful prior use of GANs in related works like AlphaSRGAN and SRDRM-GAN, we are motivated to leverage the strengths of GANs for underwater image super-resolution. GANs have shown remarkable performance in generating realistic images while improving perceptual quality. Inspired by these advancements, we adopt a state-of-the-art GAN-based super-resolution model called Real-ESRGAN in our work.

The real-ESRGAN [3] model can effectively improve the visual quality of general LR images. A high-order degradation process was used in this model that more accurately emulates complicated and realistic degradations encountered in real-world scenarios. The authors employed a U-Net based discriminator [7] with skip connections to increase discrimination power. Furthermore, they used spectral normalization [8] to stabilize the training process.

In this study, the Real-ESRGAN model is specifically adapted and optimized for the task of enhancing the visual quality of underwater images.

III. METHODOLOGY

In this section, we first present an overview of the dataset used in this paper. Then we explain the method used to generate LR images. Finally, we discuss the transfer learning technique and delve into the details of the fine-tuning process.

A. Dataset

We utilize the USR-248 dataset [1] for fine-tuning and testing the model. The USR-248 dataset is specifically used for underwater image super-resolution. In this dataset, the training and testing folders each contain multiple sets of LR images paired with the corresponding set of HR images. Each set in the training folder comprises 1060 images, and each set in the testing folder contains 248 images. The original HR images have a resolution of 640 × 480. The USR-248 dataset offers a diverse range of underwater scenes and objects of interest, including various backgrounds and subjects such as coral reefs, fish, divers, and wrecks, and provides a comprehensive collection for training and evaluation purposes.

To fine-tune the model, we use HR images from the training folder. We employ the Real-ESRGAN model's advanced degradation process to obtain LR images from the HR images. This LR image generation occurs during the fine-tuning process. The details of this degradation process will be explained in Subsection B. To assess the model's performance, we use HR and 4× downsampled images from the testing folder.

B. Degradation process

In underwater imaging, degradation factors like absorption, scattering, and poor visibility affect image quality. Classical degradation methods are insufficient for simulating these real-world underwater degradations. To address this limitation, it becomes essential to employ a suitable degradation process that better simulates real-world degradations.

The high-order degradation process employed in our research is based on the work of Wang et al. [3]. This process is highly effective for simulating realistic degradations. Employing the high-order degradation process can also be useful.
for simulating real-world underwater degradations. By integrating this degradation model, we achieve substantial improvements in preserving fine details, reducing noise, and eliminating artifacts. These advancements result in superior perceptual quality and improved performance in underwater image super-resolution tasks.

C. Transfer learning and fine-tuning details

Transfer learning technique offers several advantages in the context of model training. This technique is useful in compensating for the shortage of data and improves the performance of the model. Moreover, transfer learning facilitates faster convergence of the model and improves generalization capability to unseen data. These benefits make transfer learning a valuable technique for achieving superior performance and efficiency in various deep-learning tasks.

In our proposed approach, instead of training the Real-ESRGAN model from scratch, we use the transfer learning technique. In addition to the advantages mentioned in the previous paragraph, employing this technique enables us to save significant time and computational resources that are required to train the Real-ESRGAN model from scratch. First, we load the pre-trained RealESRGAN x4plus model. This model has been trained on natural image datasets. Next, we fine-tune this model on the USR-248 dataset of underwater images.

In this work, Adam optimizer is used for fine-tuning the pre-trained Real-ESRGAN model. We employ a combination of L1 loss, perceptual loss, and GAN loss functions for fine-tuning. The model is fine-tuned for 2200 iterations (approximately 20 epochs) with a learning rate of 0.0001. To accomplish this, we utilize Google Colab GPU with a batch size of 10.

IV. RESULTS

In this section, we provide a comparison between our fine-tuned model and the baseline Real-ESRGAN model to highlight the improvements achieved through the fine-tuning process.

Fig. 1. The columns represent the input images, the outputs of Real-ESRGAN, the outputs of fine-tuned Real-ESRGAN, and the original images, respectively.
Furthermore, we demonstrate the model’s capabilities for real-world underwater applications.

A. Evaluation method

As described in the SRGAN [9] paper, the existing quantitative metrics, such as peak signal-to-noise ratio (PSNR), lack the ability to precisely mirror the visual preferences of human observers. These metrics often fail to evaluate intricate texture information and fine details. It is important to note that achieving high PSNR value does not necessarily guarantee superior perceptual quality. To overcome these limitations, we rely on qualitative results for a more detailed evaluation of the models. Qualitative assessment provides a more accurate and comprehensive understanding of the model’s performance in underwater image super-resolution. By visually analyzing the generated images, we can assess the preservation of fine details, while also evaluating the realism and overall visual quality of the images.

In our experiments, the resolution of the input images is increased by using a scale factor of 4×. Fig. 1 exhibits the results for different images from the testing folder of the USR-248

![Image](image_url)
dataset. Furthermore, for a better comparison of fine details, magnified regions of underwater images have been shown in Fig. 2.

B. Results analysis

The obtained results demonstrate that the fine-tuned model outperforms the Real-ESRGAN model in terms of recovering realistic and natural textures and generating images with enhanced visual quality. Furthermore, the comparison reveals that the fine-tuned model can better preserve fine details, such as intricate textures. These findings confirm the effectiveness of our fine-tuning approach in achieving superior performance and producing higher-quality images in the context of underwater image super-resolution.

Our proposed model exhibits great potential for a wide range of real-world underwater applications. AUVs can benefit from this model in tasks like underwater mapping, object detection, and navigation, enabling more accurate data collection and analysis. In marine research, our model can facilitate species identification and analysis of underwater ecosystems by improving the visual quality of underwater images. Furthermore, in underwater archaeology, underwater surveillance systems, and other related fields, the enhanced image quality provided by the model can improve decision-making and analysis in challenging underwater environments.

V. CONCLUSION AND FUTURE WORK

This research paper focused on optimizing the Real-ESRGAN model with the specific objective of enhancing the visual quality of underwater images. Our proposed model outperforms the Real-ESRGAN model and achieves superior visual quality, making it well-suited for real-world underwater applications.

As a potential direction for future work, the A-ESRGAN model [10] can be optimized to improve the visual quality of underwater images. This model incorporates a multi-scale attention U-Net discriminator and demonstrates improved performance compared to the Real-ESRGAN model.

REFERENCES

[1] M. J. Islam, S. S. Enan, P. Luo, and J. Sattar, "Underwater image super-resolution using deep residual multipliers," in 2020 IEEE International Conference on Robotics and Automation (ICRA), May 2020, pp. 900-906.
[2] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, et al., "Generative adversarial nets," in NIPS, 2014.
[3] X. Wang, L. Xie, C. Dong, and Y. Shan, "Real-esrgan: Training real-world blind super-resolution with pure synthetic data," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 1905-1914.
[4] H. Lu, Y. Li, S. Nakashima, H. J. Kim, and S. Serikawa, "Underwater Image Super-Resolution by Descattering and Fusion," IEEE Access, vol. 5, pp. 670-679, 2017.
[5] A. K. Cherian and E. Poovammal, "A novel alphasrgan for underwater image super resolution," Computers, Materials & Continua, vol. 69, no. 2, pp. 1537-1552, 2021.
[6] H. Wang, H. Wu, Q. Hu, J. Chi, X. Yu, et al., "Underwater image super-resolution using multi-stage information distillation networks," J. Vis. Commun. Image Represent., vol. 77, article 103136, 2021.
[7] E. Schonfeld, B. Schiele, and A. Khoreva, "A u-net based discriminator for generative adversarial networks," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 8207-8216.
[8] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," arXiv preprint arXiv:1802.05957, Feb. 2018.
[9] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, et al., "Photo-realistic single image super-resolution using a generative adversarial network," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4681-4690.
[10] Z. Wei, Y. Huang, Y. Chen, C. Zheng, and J. Gao, "A-ESRGAN: Training Real-World Blind Super-Resolution with Attention U-Net Discriminators," arXiv preprint arXiv:2112.10046.