Enhancing Classification of Aquatic Species through Supervised Contrastive Learning and Advanced Image Super-Resolution

Sadia Nasrin Tisha
University of Maryland, Baltimore County
Baltimore, Maryland, USA
stisha1@umbc.edu

Geethanjali Nallani
University of Maryland, Baltimore County
Baltimore, Maryland, USA
le74652@umbc.edu

Abstract

Zebra and quagga mussels, invasive species in North America, cause significant ecological and economic damage by out-competing native species and obstructing water infrastructure. Traditional methods for detecting mussel larvae are costly and time-consuming, necessitating the development of automated monitoring procedures and super-resolution (SISR) techniques to enhance low-resolution water sample images. This study investigates the effectiveness of three advanced SISR methods—Super-Resolution Convolutional Neural Network (SRCNN), Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN), and Water-Net—across three classification frameworks: a baseline model, a supervised contrastive learning model, and a convolutional neural network (CNN) in improving the classification accuracy of invasive and non-invasive larvae. By integrating these SISR techniques with supervised contrastive learning, the aim was to enhance image quality and feature representation. The results demonstrate that ESRGAN significantly outperforms SRCNN and Water-Net, achieving the highest classification accuracy of 96.96% due to its advanced architecture and feature enhancement capabilities. These findings underscore the critical role of image quality in species classification and highlight the potential of ESRGAN in supporting effective ecological management and mitigation strategies for invasive species. The code of this study is available here.

1. Introduction

Zebra and Quagga mussels are native to Eurasia but have become widely introduced in North American waters, causing ecological disruption[35]. The rapid spread of these organisms leads to the formation of large colonies that restrict water flow and interfere with power generation from water systems and also causing ecological damage[5]. The annual economic impact of invasive species, specifically power plants and municipal drinking water systems in North America, has been estimated at between $267 million and $1 billion [5, 18]. Dreissenid mussels spread rapidly, laying millions of eggs annually. Once established, their population growth is uncontrollable, and monitoring their presence at the larval stage is crucial[18]. The conventional methods of detecting veliger presence are to collect plankton or water samples and then examine the selection using cross-polarized light microscopy[18] or environmental DNA[39]. Both of these methods are costly, time-consuming, and require human experts. Developing an automated procedure to monitor veliger larvae from water sample videos visually is essential to address these limitations. This project aims to classify invasive and non-invasive larvae using videos of water samples, a task complicated by the low-resolution images typically captured in such videos. This low resolution can lead to misidentification and false negatives, undermining management and control measures.

To address the challenge of low-resolution images in water sample videos, single-image super-resolution (SISR) techniques enhance image quality by recovering high-resolution images from low-resolution ones, addressing the challenge of low-resolution images in water sample videos. This problem has gained significant attention in the research community and AI industry due to its potential to enhance image quality. The use of deep learning, particularly Convolutional Neural Networks (CNNs), has transformed SISR techniques, starting with the pioneering work of Dong et al. [9] on the Super-Resolution Convolutional Neural Network (SRCNN). SRCNN's ability to directly learn the mapping between low and high-resolution image patches has made it particularly effective for enhancing fine details, which is crucial for tasks such as accurate larval identification in water samples. While SRCNN set the stage, subsequent innovations in network architecture and training strategies have continuously improved super-resolution performance, particularly regarding Peak signal-to-noise ratio (PSNR) [14, 22–24, 26, 42, 43, 58, 59]. However, methods optimized solely for PSNR often produce over-smoothed images needing more high-frequency details, deviating from the perceptual quality assessed by human observers [26]. Generative Adversarial Networks (GANs) [12] have also been incorporated to encourage networks to generate more natural-looking images. Semantic image priors have also been integrated to enhance texture detail recovery [37, 49]. One notable advancement in perceptual-driven SISR is the Super-Resolution Generative Adversarial Network (SRGAN) [29, 34], which combines residual blocks with a GAN framework optimized using perceptual loss. SRGAN significantly enhances the visual quality of reconstructed images compared to PSNR-oriented methods. Building on SRGAN, Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) [19, 50] introduce further enhancements, including deeper generator networks with residual-in-residual dense blocks and a relativistic discriminator that evaluates image realism relative to real images. These improvements enable ESRGAN to generate high-fidelity images with superior texture details, essential for distinguishing between invasive and non-invasive larvae. On the other hand, the Water-Net architecture [27] has been developed to enhance aquatic images, especially underwater imaging. It integrates domain-specific knowledge to address challenges such as low contrast and color distortions typical of underwater environments. Water-Net uses a multi-scale feature extraction mechanism and a context aggregation strategy to enhance image clarity and detail. These specialized features make Water-Net particularly effective for improving the visual quality of water sample images, facilitating more accurate identification of larval species. This study aims to validate the effectiveness of three advanced SISR techniques which offers unique advantages for enhancing image quality, especially in monitoring aquatic invasive species. By improving the resolution and clarity of water sample images, these techniques aim to significantly enhance the accuracy of classifying invasive and non-invasive larvae. This, in turn, supports more effective management strategies for mitigating the impacts of invasive species on aquatic ecosystems.

In addition, this research combines SISR techniques with supervised contrastive learning (SCL) [21] for classification. SCL aims to create a strong feature representation by positioning similar samples (i.e., samples from the same class) closer together in the feature space while positioning dissimilar samples (i.e., samples from different classes) farther apart. SCL leverages the principles of self-supervised contrastive learning within a supervised framework. The main idea involves using an anchor sample and corresponding positive and negative samples for each anchor. The distance between the anchor and negative samples is maximized in the embedding space, while the distance between the anchor and positive samples is minimized. Positive samples are typically created through data augmentation of the anchor, whereas negative samples are selected from other examples within the batch. In contrast to traditional supervised learning, contrastive learning captures the intrinsic structure of the data [44], enhancing the model's ability to generalize and perform well on various downstream tasks, especially in scenarios with limited labeled data [13, 17, 25, 32, 48, 54]. Contrastive learning has proven effective in computer vision tasks, particularly in enhancing feature representation. Using SISR-enhanced images, the classification model can better distinguish between invasive and non-invasive larvae, crucial for managing invasive species in aquatic ecosystems.

The main research question of this study is: "How does enhancing image quality using super-resolution techniques affect the performance of supervised contrastive learning models in classifying aquatic species?" This question is crucial as it addresses a significant gap in understanding the relationship between image quality and species classification accuracy. To investigate this question, the study will utilize advanced SISR techniques to improve the resolution of water sample images. These enhanced images will then train and evaluate supervised contrastive learning models, baseline models, and basic CNN models. Figure 1 shows the workflow of this study. The study will use a carefully selected dataset containing images of invasive and non-invasive aquatic species captured from water samples via video footage. By thoroughly assessing the impact of image quality enhancement on classification performance, this research aims to offer valuable insights into the effectiveness of these techniques for monitoring and managing invasive species in aquatic ecosystems. The study's findings are expected to contribute to developing more accurate and efficient methods for detecting and classifying invasive Dreissenid mussels, thereby supporting improved ecological management and mitigation strategies. Preliminary results suggest that the contrastive learning approach significantly improves the classification of invasive species, and the application of super-resolution techniques further enhances model accuracy through robust feature selection. This research emphasizes the significance of image quality in the classification process and demonstrates the potential of advanced super-resolution and machine-learning techniques in addressing critical environmental challenges.

2. Related Works

2.1. The Ecological and Economic Imperative for Advanced Classification of Invasive Species

Invasive species like quagga and zebra mussels have disrupted the United States' complex ecosystems. Due to their widespread expansion, many states—most notably Texas, California, and Arizona—have implemented exten-

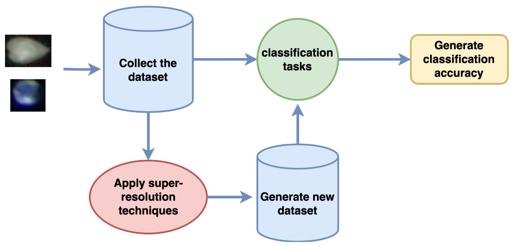


Figure 1. The workflow of this study

sive monitoring and control programs designed to manage the sneaky growth of these organisms [5, 31]. The growing understanding of such species' significant ecological and economic consequences drives this response [5, 16, 31]. Biological invaders cost around $219 billion domestically in the United States and perhaps exceed $4 trillion internationally [5]. The economic and environmental impact highlights the urgent need for successful and efficient containment techniques. In light of this, automated early detection systems have emerged as a promising strategy to curb the spread of invasive mussels [5].

Traditional methods like environmental DNA (eDNA) analysis and microscope photography have proven effective, but they demand significant financial and time resources and specialized expertise for accurate identification [18]. To overcome these challenges, deep learning for image identification in aquatic environments has gained popularity, which offers a potent solution for analyzing visual data to detect invasive species [3, 4, 15]. Our research is focused on identifying invasive species using video samples collected from the Colorado River near the Davis Dam in Arizona [47]. We are highlighting the effectiveness of advanced detection techniques in this context. Given the limitations of traditional methods, our study explores using contrastive learning techniques to improve classification efficiency. Contrastive learning aims to bring similar data points closer and push dissimilar points further apart. This is done by optimizing a contrastive loss function, and a classifier can be added for supervised image classification.

Contrastive learning is particularly effective at learning from limited labeled data and has demonstrated success in various downstream tasks in computer vision [8, 25, 28, 32, 40, 48, 54]. Here, Diba et al. [8] have introduced a novel self-supervised learning method called Vi2CLR, which processes images and videos simultaneously to learn solid visual representations. The Vi2CLR method utilizes a neural network architecture that leverages unlabeled videos to capture dynamic and static visual cues, enabling instance similarity/dissimilarity learning through visual clustering. By integrating 2D (image) and 3D (video) Convolutional Neural Networks (CNNs), Vi2CLR achieves performance comparable to supervised learning. Extensive evaluations of various datasets have demonstrated that Vi2CLR outperforms state-of-the-art self-supervised methods in action recognition, image classification, and object classification. We utilize the supervised contrastive learning framework introduced by Khosla et al. [21], where the encoder learns a discriminative representation of the data, enhancing the classification of invasive species in aquatic environments. This approach aims to improve detection accuracy and efficiency, contributing to more effective management and mitigation strategies for invasive Dreissenid mussels.

2.2. Addressing Low-Resolution Challenges in Aquatic Species Identification Using SR Techniques

In this study, we face a significant challenge: low-resolution images in water sample videos. This issue is of utmost importance as it hampers our ability to identify and classify aquatic species accurately. We can use super-resolution (SR) [36] techniques to recover high-resolution (HR) images from low-resolution (LR) ones. This involves restoring HR images from one or more LR observations of the same scene. Depending on the number of input LR images, SR can be categorized into Single Image Super-Resolution (SISR) and Multi-Image Super-Resolution (MISR). SISR, in particular, is highly efficient compared to MISR. HR images with high perceptual quality are invaluable in various medical, satellite, and security imaging domains. Mainstream SISR algorithms are categorized as interpolation-based, reconstruction-based, and learning-based methods.

Interpolation-based methods, such as bicubic interpolation [20] and Lanczos resampling [10], are known for their speed and simplicity but often suffer from accuracy limitations. Reconstruction-based methods [6, 33, 41, 51] employ sophisticated prior knowledge to constrain the solution space, allowing for the generation of flexible and sharp details. However, their performance typically degrades as the scale factor increases and is generally time-consuming. Learning-based methods, also called example-based methods, have gained prominence due to their fast computation and superior performance. These methods utilize machine learning algorithms to analyze statistical relationships between LR and HR images from extensive training examples. The Markov Random Field (MRF) approach was first introduced by Freeman et al. [11] to synthesize visually pleasing image textures using abundant real-world images. Neighbor embedding methods proposed by Chang et al. [2] leveraged similar local geometries between LR and HR patches to restore HR images. Researchers were inspired by sparse signal recovery theory [1] and applied sparse coding methods [30, 45, 46, 52, 56] to SISR problems. Random forest techniques [38] have recently been used to enhance reconstruction performance. Additionally, several works have combined the advantages of reconstruction-based methods with learning-based approaches to reduce further artifacts introduced by external training examples [7, 53, 55, 57].

Deep learning (DL)–based SISR algorithms have recently significantly improved over traditional methods. This study uses advanced SISR techniques to enhance the resolution of water sample images. SRCNN utilizes a deep convolutional neural network to learn the mapping between LR and HR images, effectively enhancing fine details important for accurate species identification. ESRGAN, based on the GAN framework, introduces enhancements like deeper networks with residual-in-residual dense blocks and a relativistic discriminator to produce high-fidelity images with superior texture details. The Water-Net architecture, designed specifically for aquatic image enhancement, incorporates domain-specific knowledge to address challenges such as low contrast and color distortions in underwater imaging. By leveraging these advanced SISR techniques, this study aims to significantly improve the image quality of water sample videos, thereby enhancing the classification of invasive and non-invasive aquatic species.

3. Invasive Species Dataset

This study uses high-definition video recordings of aquatic streams to create a specialized dataset. A Kalman Filter-based proprietary algorithm was used to identify, track, and extract larvae imagery from the videos, providing multiple frame sequences for each organism. The extracted images are then classified as either non-invasive or invasive. Figure 2 displays examples of these image sets. We characterized the dataset by categorizing its contents into two distinct classes: invasive and non-invasive species. Specifically, it encompasses 6,905 individual organisms, segmented into 1,220 invasive and 5,685 non-invasive. The dataset comprises 88,050 images, split into 44,646 images of invasive species and 43,404 images of non-invasive species. We represented each organism within the dataset by a sequence of frames, varying in number from a minimum of 6 to a maximum of 42 per organism.

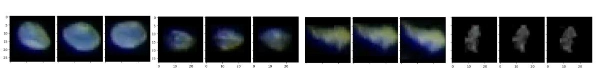


Figure 2. An example of invasive dreissenid and non-invasive species in our dataset. The image has four different organisms, with the first six columns containing images of two different invasive larvae and the following six columns containing two different non-invasive larvae.

4. Methodology

4.1. Supervised Contrastive Learning

Our training methodology is based on the framework established by Khosla et al. [21] in their supervised contrastive learning study. The development of the classification model involves two primary steps: (1) training an encoder to generate representation vectors from input images in such a way that vectors from the same class are similar while vectors from different classes are distinct, and (2) training a classifier on top of the frozen encoder.

Given a batch of input data, we apply data augmentation twice to each image and feed the augmented images to the encoder network. For an input image x, we create $\hat{x} = \mathrm{Aug}(x)$, where Aug() is a function that applies various data augmentations, such as random flips and rotations. The encoder then generates a representation vector $r = \mathrm{Enc}(x)$, where $r \in \mathbb{R}^d$. In our case, we use the ResNet50V2 architecture to build the encoder. A projection network maps r to z where the contrastive loss is applied, with $z = \mathrm{Proj}(r)$.

When labels are not provided, the loss is calculated using the augmented sample as a positive example. The loss function for a sample in the batch is defined as:

$$L = -\log \frac{\exp(\mathrm{sim}(z_i, z_j)/\tau)}{\sum_{k \in A(i)} \exp(\mathrm{sim}(z_i, z_k)/\tau)}$$

Here, z_i is the anchor, z_j is the augmented positive sample, and z_k are the negative samples in a mini-batch. The function sim() calculates the cosine similarity between two vectors, and τ is a scalar temperature parameter. Each anchor has one positive example and $2(N-1)$ negative examples. $A(i)$ has a total of $2N-1$ terms and is defined as:

$$A(i) = \{k | k \in \{1, 2, \ldots, 2N\}, k \neq i\}$$

In supervised contrastive learning, there are more positive samples than just the augmented image. To incorporate label information, we modify the loss function as follows:

$$L = -\frac{1}{P(i)} \sum_{p \in P(i)} \log \frac{\exp(\mathrm{sim}(z_i, z_p)/\tau)}{\sum_{k \in A(i)} \exp(\mathrm{sim}(z_i, z_k)/\tau)}$$

Here, $P(i)$ includes indices of all positive samples in the batch, including the augmented image. Supervised contrastive loss performs better with a large batch size due to more negative samples.

After training the encoder with contrastive loss, we freeze it and add a classifier. The classifier is then trained to make the final predictions using cross-entropy loss. This two-step process ensures that the learned representations are highly discriminative, enabling accurate classification of aquatic species based on enhanced water sample images.

4.2. Super-Resolution Techniques

4.2.1 Super-Resolution Convolutional Neural Network (SRCNN)

The SRCNN is a method used to enhance image resolution. It involves a multi-step process that begins with upscaling LR images using bicubic interpolation. A three-layer convolutional neural network then processes the upscaled image. The first layer, the patch extraction and representation layer, extracts overlapping patches from the upscaled image and transforms them into high-dimensional vectors through convolutional filtering. The second layer, the nonlinear mapping layer, applies complex transformations to these vectors, learning the intricate relationships between LR and HR features. Finally, the reconstruction layer synthesizes the HR image from these transformed vectors, enhancing image resolution. This process involves successive convolution operations with learned weights and biases optimized through training on large datasets of LR-HR image pairs. The goal of employing SRCNN is to significantly improve the resolution of water sample images, enabling more accurate classification of aquatic species.

4.2.2 Enhanced Super-Resolution Generative Adversarial Network (ESRGAN)

The ESRGAN is an advanced model that enhances image quality by generating high-resolution outputs from low-resolution inputs. It utilizes multiple convolutional layers to extract detailed features from the low-resolution image. The model also incorporates Residual-in-Residual Dense Blocks (RRDB) to improve feature extraction and mitigate the vanishing gradient problem without relying on batch normalization. Sub-pixel convolutional layers are then used to upscale the image resolution, resulting in a final high-resolution image with enhanced details. This advanced architecture significantly improves the resolution and quality of water sample images, facilitating more accurate classification of aquatic species.

4.2.3 Water-Net

Water-Net is a specialized neural network designed to improve water quality assessment from images. It uses a gated fusion approach, starting by transforming input images through Feature Transformation Units (FTUs). These units apply convolutional operations to extract and refine high-dimensional feature maps. The refined features are then used to predict confidence maps, representing the network's certainty in various aspects of the image. In the final stage, a gated fusion mechanism combines the refined features and corresponding confidence maps, integrating spatial and confidence information to produce an enhanced image. This advanced architecture significantly enhances image quality, aiding in accurately classifying aquatic species and improving water quality monitoring and assessment.

5. Result Analysis

5.1. Super-Resolution Techniques

In this study, we are assessing the performance of SR techniques on our underwater invasive species image dataset and comparing the results with real-world images. Here, table 1 displays the invasive species data, while table 2 illustrates real-world images. Each image was processed using three super-resolution methods: SRCNN, ESRGAN, and Water-Net. We used evaluation metrics such as Peak Signal-to-Noise Ratio (PSNR), Mean Squared Error (MSE), and Structural Similarity Index Measure (SSIM) to compare the original images, degraded images, and the super-resolution (SR) images produced by each method. We then analyzed the qualitative and quantitative results for each method.

5.1.1 SRCNN

Quantitative Perspective:
In the case of real-world data, the SRCNN method showed an improvement in image quality based on several key metrics. The PSNR increased from 34.46 (degraded) to 34.73 (SR), while the MSE decreased from 69.85 (degraded) to 65.68 (SR). Additionally, the SSIM improved from 0.92 (degraded) to 0.93 (SR). Compared to the original image scores of PSNR: 30.507, MSE: 66.548, and SSIM: 0.926, it is evident that SRCNN effectively enhances real-world images. These metrics indicate that SRCNN effectively reduces reconstruction error (lower MSE) and increases the perceptual quality and structural similarity (higher PSNR and SSIM) of the super-resolved images relative to the original degraded images. Conversely, the performance of the SRCNN method on the invasive species dataset posed significant challenges. Here, the PSNR experienced a slight decrease from 31.547 (degraded) to 30.53 (SR), MSE increased from 225.14 (degraded) to 358.48 (SR), and SSIM showed a marginal increase from 0.87 (degraded) to 0.902 (SR). When compared to the original image scores, it becomes apparent that while the SSIM marginally improved, the PSNR decreased, and MSE increased significantly. These findings suggest that SRCNN encounters difficulties in enhancing underwater microscopy images, as evidenced by the higher reconstruction error and lower perceptual quality relative to the original image.
Qualitative Perspective:
The SRCNN method demonstrated significant improvements in real-world data. The super-resolved (SR) images displayed enhanced clarity and detail, resembling the original high-resolution images. This visual enhancement is supported by quantitative improvements in PSNR, MSE,

and SSIM, indicating the effectiveness of SRCNN in improving real-world image quality. However, the SR images produced by SRCNN did not exhibit significant visual enhancement for the invasive species dataset. These images often appeared more blurred and needed more fine details to identify invasive species accurately. This visual outcome aligns with the quantitative metrics, which showed increased MSE and decreased PSNR, suggesting poorer image reconstruction and perceptual quality.

5.1.2 ESRGAN

Quantitative Perspective:
ESRGAN has shown outstanding performance with both real-world and invasive species data. In the case of the real-world image, ESRGAN significantly improved the PSNR from 30.72 (degraded) to 38.93 (SR), reduced the MSE from 54.98 (degraded) to 8.31 (SR), and enhanced the SSIM from 0.82 (degraded) to 0.97 (SR). When compared to the original image scores of PSNR: 30.507, MSE: 66.548, and SSIM: 0.926, it is clear that ESRGAN not only surpasses the degraded image quality but also exceeds the quality metrics of the original images. Similarly, for the invasive species dataset, ESRGAN significantly increased the PSNR from 24.27 (degraded) to 36.89 (SR), reduced the MSE from 242.84 (degraded) to 13.28 (SR), and improved the SSIM from 0.72 (degraded) to 0.98 (SR). These improvements are also substantial when compared to the original image scores, demonstrating ESRGAN's ability to greatly enhance underwater microscopy images, surpassing the original benchmarks in quality.

Qualitative Perspective:
ESRGAN produced impressive results visually for both sets of data. The super-resolved (SR) images showed high fidelity and detailed textures, significantly improving sharpness and natural appearance compared to the degraded images. In the case of real-world data, the SR images were almost identical to high-quality originals, displaying clear and detailed features. In the invasive species dataset, ESRGAN effectively highlighted intricate textures and details crucial for species identification.

5.1.3 Water-Net

Quantitative Perspective:
In the invasive species dataset, Water-Net improved the PSNR from 16.29 (degraded) to 27.74 (SR), decreased MSE from 1527.6 (degraded) to 109.27 (SR), and increased SSIM from 0.12 (degraded) to 0.65 (SR). These improvements are significant when compared to the original image scores. While Water-Net does not surpass the original image quality metrics in PSNR and SSIM, the enhancements over the degraded images are considerable, especially in SSIM, which is crucial for structural similarity. For the real-world data, Water-Net presented mixed quantitative results. The PSNR increased significantly from 3.25 (degraded) to 27.05 (SR), indicating a substantial enhancement in image quality. The MSE showed a dramatic decrease from 3072.6 (degraded) to 128.17 (SR), highlighting a reduction in reconstruction error. The SSIM increased dramatically from 0.004 (degraded) to 0.96 (SR), indicating a significant improvement in structural similarity to the original image. Compared to the original image scores, it is clear that Water-Net, while improving SSIM significantly, still lags behind the original PSNR and MSE values but substantially enhances the degraded images.

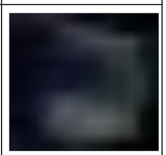
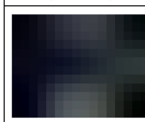
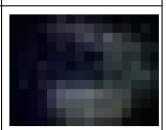
Qualitative Perspective:
Water-Net displayed significant improvements in the real world data, especially in SSIM, indicating better preservation of structural details and overall image quality closer to the original. The enhanced images clearly showed distinct features previously blurred in the degraded images, aligning with the notable increase in SSIM. Water-Net effectively improved visual clarity and structural details for the invasive species dataset. Although the quantitative metrics, such as PSNR and MSE, were not the highest compared to other methods, the visual improvements were remarkable. The super-resolution images exhibited better-defined features and textures necessary for accurate species identification, which is crucial for underwater images.

5.2. Classification Task Analysis

After applying super-resolution techniques to the invasive species dataset, we created new datasets for each super-resolution method while retaining the original dataset. Subsequently, we performed classification tasks to distinguish between invasive and non-invasive species across these new datasets. The table 3 compares the performance of three super-resolution techniques—SRCNN, ESRGAN, and Water-Net—alongside the standard resolution (original dataset) in this classification task. This evaluation used three models: a baseline classification model, a supervised contrastive learning model, and a convolutional neural network (CNN).

Our findings demonstrate the superiority of ESRGAN in enhancing images for classification tasks. The baseline classification model, which jointly optimized both the encoder and the classifier to minimize cross-entropy loss, showed that ESRGAN-enhanced images significantly improved accuracy. In contrast, SRCNN and Water-Net performed better than the standard resolution dataset, suggesting that these methods may not effectively enhance the key points. The Supervised Contrastive Learning Model was trained in two phases optimizing the fully connected layers with softmax activation. The ESRGAN-enhanced images showed a notable improvement, achieving the highest accuracy of 96.96%. This indicates that ESRGAN is particularly effective in enhancing features crucial for classification. The accuracy for standard resolution images slightly decreased compared to the baseline model. Although SRCNN and Water-Net also performed worse than standard resolution images, the drop in performance was less severe than in the baseline model, highlighting the robustness of the supervised contrastive learning approach. In addition, the CNN model comprised two convolutional layers followed by max pooling, a flattened layer, and two dense layers. For binary classification, the final output layer used softmax activation. The model was compiled with the Adam optimizer and categorical cross-entropy loss. Once again, ESRGAN-enhanced images exhibited the highest accuracy. Standard resolution images followed, with Water-Net and SRCNN-enhanced images showing the lowest performance.

Table 1. Theevaluation of SR techniques for invasive species Data

		PSNR	MSE	SSIM
Original Image		PSNR: 30.507	MSE: 173.54	SSIM: 0.905
SRCNN – Degraded Image		PSNR: 31.547	MSE: 225.14	SSIM: 0.87
SRCNN – SR Image		PSNR: 30.53	MSE: 358.48	SSIM: 0.902
ESRGAN – Degraded Image		PSNR: 24.27	MSE: 242.84	SSIM: 0.72
ESRGAN – SR Image		PSNR: 36.89	MSE: 13.28	SSIM: 0.98
Water-Net – Degraded Image		PSNR: 16.29	MSE: 1527.6	SSIM: 0.12
Water-Net – SR Image		PSNR: 27.74	MSE: 109.27	SSIM: 0.65

6. Discussion

The main research question of this study is: "How does improving image quality through super-resolution techniques impact classification performance?" We also aimed to determine whether the enhancement in classification performance is primarily attributed to the SR techniques or the superior architecture of the classification models.

ESRGAN has shown superior performance due to its advanced architecture, which includes Residual-in-Residual Dense Blocks (RRDB) and a relativistic discriminator. The RRDBs capture complex patterns and delicate details by enhancing gradient flow and preserving features through multiple residual connections. The relativistic discriminator improves adversarial training by assessing the realism of generated images relative to actual photos, resulting in more natural outputs. ESRGAN achieved the highest accuracy of 96.96% in the supervised contrastive learning model, significantly outperforming other methods and standard-resolution images. On the other hand, SRCNN demonstrated limited effectiveness for underwater microscopy images due to its design, which is optimized for high-contrast natural images. The network needed help with intricate textures and lower contrast typical of underwater images, resulting in lower classification accuracy. The highest accuracy achieved with SRCNN was 91.34%, lower than the 96.29% obtained with standard resolution images, indicat-

Table 2. The evaluation of SR techniques for real world Data

		PSNR	MSE	SSIM
Original Image		PSNR: 30.507	MSE: 66.548	SSIM: 0.926
SRCNN – Degraded Image		PSNR: 34.46	MSE: 69.85	SSIM: 0.92
SRCNN – SR Image		PSNR: 34.73	MSE: 65.68	SSIM: 0.93
ESRGAN – Degraded Image		PSNR: 30.72	MSE: 54.98	SSIM: 0.82
ESRGAN – SR Image		PSNR: 38.93	MSE: 8.31	SSIM: 0.97
Water-Net – Degraded Image		PSNR: 3.25	MSE: 3072.6	SSIM: 0.004
Water-Net – SR Image		PSNR: 27.05	MSE: 128.17	SSIM: 0.96

Table 3. The performance of three super-resolution techniques (SRCNN, ESRGAN, and Water-Net) and the standard resolution images on a classification task using three different models

Model	Standard resolution	SRCNN	ESRGAN	Water-Net
Baseline classification model	**96.29%**	91.34%	**95.28%**	88.71%
Supervised contrastive learning	89.43%	87.63%	**96.96%**	88.19%
Convolutional Neural Network	86.00%	81.56%	**91.06%**	89.39%

ing its inability to enhance critical features necessary for accurate classification in complex conditions. Furthermore, Water-Net is designed specifically for enhancing underwater images, addressing issues like low contrast and color distortions. While it improves visual quality and structural similarity, it doesn't optimize quantitative metrics like PSNR and MSE as effectively as ESRGAN. Water-Net achieved an accuracy of 88.71%, surpassing SRCNN but still falling short of SR images and ESRGAN.

The study's results show that using super-resolution techniques to improve image quality significantly impacts classification performance. ESRGAN, with its advanced architectural features, demonstrates the most substantial improvement, especially in the supervised contrastive learning model. This indicates that ESRGAN's superior architecture is crucial in enhancing classification accuracy. On the other hand, SRCNN is effective for some real-world images but needs help with the complexities of underwater images, resulting in lower classification performance. Water-Net, explicitly designed for underwater image enhancement, improves visual quality and structural similarity but does not achieve the highest quantitative metrics. These findings highlight the importance of selecting super-resolution techniques based on the specific dataset and classification task. ESRGAN's advanced architecture and robust feature enhancement make it the most suitable choice for improving classification performance across diverse image datasets.

7. Limitations & Future Works

This study has identified several vital limitations that impact the performance and generalizability of the super-resolution models used. Firstly, the compatibility of the training data with the super-resolution models is crucial. The effectiveness of SRCNN, ESRGAN, and WaterNet heavily depends on the quality and diversity of the training datasets. Inconsistent or biased data can lead to suboptimal performance, mainly when applied to complex datasets such as underwater microscopy images. Secondly, super-resolution techniques can introduce artifacts that negatively affect the classification task by obscuring important features or adding noise, leading to misclassification. This sensitivity to artifacts underscores the need for robust artifact mitigation strategies within SR models. Lastly, the models exhibit varying sensitivity to changes in image quality. While ESRGAN performs well across different datasets, SRCNN and Water-Net show significant performance variability based on image characteristics, limiting their generalizability to diverse imaging conditions and datasets.

Future research should focus on several key areas to address the limitations identified in this study. Firstly, we should work on enhancing image quality by refining super-resolution algorithms and incorporating more sophisticated techniques. This will help improve image reconstruction, reduce artifacts, and enhance feature clarity. Secondly, we need to improve model training methodologies by using more significant and diverse datasets, employing advanced training strategies, and optimizing hyperparameters to enhance performance significantly. Thirdly, we should consider employing data augmentation techniques to increase the variability and robustness of training datasets, allowing models to generalize better across various scenarios. Additionally, we should explore hybrid super-resolution techniques that combine the strengths of different SR methods, such as integrating SRCNN, ESRGAN, and Water-Net into a unified framework to capture a broader range of image features and improve overall image quality. By testing models on various datasets beyond the original training data, cross-domain validation can ensure robustness and generalizability, identifying potential weaknesses and areas for improvement. Finally, investigating alternative machine learning approaches, such as unsupervised or semi-supervised learning, and expanding the range of datasets used for training and testing can provide new insights and help develop more versatile and robust models.

8. Conclusion

The study investigated how super-resolution techniques affect invasive and non-invasive species classification in underwater images. Three super-resolution models—SRCNN, ESRGAN, and Water-Net—were evaluated across three classification frameworks: a baseline model, a supervised contrastive learning model, and CNN model. ESRGAN consistently outperformed the other models, achieving the highest classification accuracy of 96.96% in the supervised contrastive learning framework. This highlights its robust architecture, which includes Residual-in-Residual Dense Blocks and a relativistic discriminator. While SRCNN is effective for natural images, it struggled with the complex textures and lower contrast of underwater images. Water-Net, explicitly designed for underwater image enhancement, showed improvements in structural similarity and visual quality but lagged in quantitative metrics compared to ESRGAN. The study's limitations include dependence on the quality and diversity of training datasets, sensitivity to super-resolution artifacts, and variability in model performance based on image quality changes. Future work should focus on enhancing super-resolution algorithms, improving model training methodologies, employing data augmentation, exploring hybrid super-resolution techniques, validating across multiple domains, and investigating alternative machine-learning approaches. These steps will help refine the application of super-resolution techniques to improve classification accuracy in diverse and complex imaging conditions.

9. Author Contributions

Task	Contribution
Project Idea Generation and Project Proposal writing	Sadia and Geethanjali
Identifying the dataset and data preprocessing	Sadia
Apply SRCNN techniques	Sadia and Geethanjali
Create New dataset generated by SRCNN	Sadia
Apply ESRGAN technique	Geethanjali
Create New dataset generated by ESRGAN	Sadia
Apply Water-Net technique	Geethanjali
Create New dataset generated by Waternet	Sadia
Apply Supervised Contrastive Learning in all datasets	Sadia
Apply other classification tasks in all the dataset	Sadia
Evaluate the results	Sadia and Geethanjali
Final report and presentation	Sadia and Geethanjali

Acknowledgement
The code of this study is available here: https://github.com/Sadia16101101/Enhancing-Classification-of-Aquatic-Species-and-Advanced-Image-Super-Resolution

References

[1] Michal Aharon, Michael Elad, and Alfred Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on signal processing*, 54(11):4311–4322, 2006. 3

[2] Hong Chang, Dit-Yan Yeung, and Yimin Xiong. Super-resolution through neighbor embedding. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, pages I–I. IEEE, 2004. 3

[3] Shaif Chowdhury and Greg Hamerly. Recognition of aquatic invasive species larvae using autoencoder-based feature averaging. In *International Symposium on Visual Computing*, pages 145–161. Springer, 2022. 3

[4] Shaif Chowdhury, Sadia Nasrin Tisha, Monica McGarrity, and Greg Hamerly. Video-based recognition of aquatic invasive species larvae using attention-lstm transformer. Technical report, EasyChair, 2023. 3

[5] Nancy A Connelly, Charles R O'Neill, Barbara A Knuth, and Tommy L Brown. Economic impacts of zebra mussels on drinking water treatment and electric power generation facilities. *Environmental management*, 40:105–112, 2007. 1, 3

[6] Shengyang Dai, Mei Han, Wei Xu, Ying Wu, Yihong Gong, and Aggelos K Katsaggelos. Softcuts: a soft edge smoothness prior for color image super-resolution. *IEEE transactions on image processing*, 18(5):969–981, 2009. 3

[7] Cheng Deng, Jie Xu, Kaibing Zhang, Dacheng Tao, Xinbo Gao, and Xuelong Li. Similarity constraints-based structured output regression machine: An approach to image super-resolution. *IEEE transactions on neural networks and learning systems*, 27(12):2472–2485, 2015. 4

[8] Ali Diba, Vivek Sharma, Reza Safdari, Dariush Lotfi, Saquib Sarfraz, Rainer Stiefelhagen, and Luc Van Gool. Vi2clr: Video and image for visual contrastive learning of representation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1502–1512, 2021. 3

[9] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13*, pages 184–199. Springer, 2014. 1

[10] Claude E Duchon. Lanczos filtering in one and two dimensions. *Journal of Applied Meteorology and Climatology*, 18 (8):1016–1022, 1979. 3

[11] William T Freeman, Thouis R Jones, and Egon C Pasztor. Example-based super-resolution. *IEEE Computer graphics and Applications*, 22(2):56–65, 2002. 3

[12] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014. 2

[13] Hongfeng Han, Nanyi Fei, Zhiwu Lu, and Ji-Rong Wen. Supervised contrastive learning for few-shot action classification. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 512–528. Springer, 2022. 2

[14] Muhammad Haris, Gregory Shakhnarovich, and Norimichi Ukita. Deep back-projection networks for super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1664–1673, 2018. 1

[15] Whitney Hassett, Julie Zimmerman, Gretchen Rollwagen-Bollens, Stephen M Bollens, and Timothy D Counihan. An experimental evaluation of the efficacy of imaging flow cytometry (flowcam) for detecting invasive dreissened and corbiculid bivalve veligers. *Lake and Reservoir Management*, 37(4):406–417, 2021. 3

[16] Amy Lee Hetherington, Lars G Rudstam, Rebecca L Schneider, Kristen T Holeck, Christopher W Hotaling, John E Cooper, and James R Jackson. Invader invaded: population dynamics of zebra mussels (dreissena polymorpha) and quagga mussels (dreissena rostriformis bugensis) in polymictic oneida lake, ny, usa (1992–2013). *Biological Invasions*, 21:1529–1544, 2019. 3

[17] Xinrong Hu, Dewen Zeng, Xiaowei Xu, and Yiyu Shi. Semi-supervised contrastive learning for label-efficient medical image segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part II 24*, pages 481–490. Springer, 2021. 2

[18] Ladd E Johnson. Enhanced early detection and enumeration of zebra mussel (dreissena spp.) veligers using cross-polarized light microscopy. *Hydrobiologia*, 312, 1995. 1, 3

[19] Xu Kang, Liang Liu, and Huadong Ma. Esr-gan: Environmental signal reconstruction learning with generative adversarial network. *IEEE Internet of Things Journal*, 8(1):636–646, 2020. 2

[20] Robert Keys. Cubic convolution interpolation for digital image processing. *IEEE transactions on acoustics, speech, and signal processing*, 29(6):1153–1160, 1981. 3

[21] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. *Advances in neural information processing systems*, 33:18661–18673, 2020. 2, 3, 4

[22] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1646–1654, 2016. 1

[23] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Deeply-recursive convolutional network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1637–1645, 2016.

[24] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 624–632, 2017. 1

[25] Phuc H Le-Khac, Graham Healy, and Alan F Smeaton. Contrastive representation learning: A framework and review. *Ieee Access*, 8:193907–193934, 2020. 2, 3

[26] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, 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*, pages 4681–4690, 2017. 1, 2

[27] Chongyi Li, Chunle Guo, Wenqi Ren, Runmin Cong, Junhui Hou, Sam Kwong, and Dacheng Tao. An underwater image enhancement benchmark dataset and beyond. *IEEE transactions on image processing*, 29:4376–4389, 2019. 2

[28] Rui Li, Yiheng Zhang, Zhaofan Qiu, Ting Yao, Dong Liu, and Tao Mei. Motion-focused contrastive learning of video representations. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 2105–2114, 2021. 3

[29] Baozhong Liu and Ji Chen. A super resolution algorithm based on attention mechanism and srgan network. *IEEE Access*, 9:139138–139145, 2021. 2

[30] Jiaying Liu, Wenhan Yang, Xinfeng Zhang, and Zongming Guo. Retrieval compensated group structured sparsity for image super-resolution. *IEEE Transactions on Multimedia*, 19(2):302–316, 2016. 3

[31] Frances Lucy and Elizabeth Muckle-Jeffs. History of the zebra mussel/icais conference series. *Aquatic Invasions*, 2010. 3

[32] Orchid Majumder, Avinash Ravichandran, Subhransu Maji, Alessandro Achille, Marzia Polito, and Stefano Soatto. Supervised momentum contrastive learning for few-shot classification. *arXiv preprint arXiv:2101.11058*, 2021. 2, 3

[33] Antonio Marquina and Stanley J Osher. Image super-resolution by tv-regularization and bregman iteration. *Journal of Scientific Computing*, 37:367–382, 2008. 3

[34] Yudai Nagano and Yohei Kikuta. Srgan for super-resolving low-resolution food images. In *Proceedings of the Joint Workshop on Multimedia for Cooking and Eating Activities and Multimedia Assisted Dietary Management*, pages 33–37, 2018. 2

[35] Thomas F Nalepa and Don W Schloesser. *Quagga and zebra mussels: biology, impacts, and control*. CRC press, 2013. 1

[36] Sung Cheol Park, Min Kyu Park, and Moon Gi Kang. Super-resolution image reconstruction: a technical overview. *IEEE signal processing magazine*, 20(3):21–36, 2003. 3

[37] Mehdi SM Sajjadi, Bernhard Scholkopf, and Michael Hirsch. Enhancenet: Single image super-resolution through automated texture synthesis. In *Proceedings of the IEEE international conference on computer vision*, pages 4491–4500, 2017. 2

[38] Samuel Schulter, Christian Leistner, and Horst Bischof. Fast and accurate image upscaling with super-resolution forests. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3791–3799, 2015. 3

[39] Adam J Sepulveda, Jon J Amberg, and Erik Hanson. Using environmental dna to extend the window of early detection for dreissenid mussels. *Management of Biological Invasions*, 10(2), 2019. 1

[40] Pierre Sermanet, Corey Lynch, Jasmine Hsu, and Sergey Levine. Time-contrastive networks: Self-supervised learning from multi-view observation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 486–487. IEEE, 2017. 3

[41] Jian Sun, Zongben Xu, and Heung-Yeung Shum. Image super-resolution using gradient profile prior. In *2008 IEEE conference on computer vision and pattern recognition*, pages 1–8. IEEE, 2008. 3

[42] Ying Tai, Jian Yang, and Xiaoming Liu. Image super-resolution via deep recursive residual network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3147–3155, 2017. 1

[43] Ying Tai, Jian Yang, Xiaoming Liu, and Chunyan Xu. Memnet: A persistent memory network for image restoration. In *Proceedings of the IEEE international conference on computer vision*, pages 4539–4547, 2017. 1

[44] Yonglong Tian, Chen Sun, Ben Poole, Dilip Krishnan, Cordelia Schmid, and Phillip Isola. What makes for good views for contrastive learning? *Advances in neural information processing systems*, 33:6827–6839, 2020. 2

[45] Radu Timofte, Vincent De Smet, and Luc Van Gool. Anchored neighborhood regression for fast example-based super-resolution. In *Proceedings of the IEEE international conference on computer vision*, pages 1920–1927, 2013. 3

[46] Radu Timofte, Vincent De Smet, and Luc Van Gool. A+: Adjusted anchored neighborhood regression for fast super-resolution. In *Computer Vision–ACCV 2014: 12th Asian Conference on Computer Vision, Singapore, Singapore, November 1-5, 2014, Revised Selected Papers, Part IV 12*, pages 111–126. Springer, 2015. 3

[47] Kent Turner, Wai Hing Wong, Shawn Gerstenberger, and Jennell M Miller. Interagency monitoring action plan (i-map) for quagga mussels in lake mead, nevada-arizona, usa. *Aquatic Invasions*, 6(2):195, 2011. 3

[48] Peng Wang, Kai Han, Xiu-Shen Wei, Lei Zhang, and Lei Wang. Contrastive learning based hybrid networks for long-tailed image classification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 943–952, 2021. 2, 3

[49] Xintao Wang, Ke Yu, Chao Dong, and Chen Change Loy. Recovering realistic texture in image super-resolution by deep spatial feature transform. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 606–615, 2018. 2

[50] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops*, pages 0–0, 2018. 2

[51] Qing Yan, Yi Xu, Xiaokang Yang, and Truong Q Nguyen. Single image superresolution based on gradient profile sharpness. *IEEE Transactions on Image Processing*, 24(10): 3187–3202, 2015. 3

[52] Jianchao Yang, John Wright, Thomas S Huang, and Yi Ma. Image super-resolution via sparse representation. *IEEE transactions on image processing*, 19(11):2861–2873, 2010. 3

[53] Wenming Yang, Yapeng Tian, Fei Zhou, Qingmin Liao, Hai Chen, and Chenglin Zheng. Consistent coding scheme for single-image super-resolution via independent dictionaries. *IEEE Transactions on Multimedia*, 18(3):313–325, 2016. 4

[54] Zhanyuan Yang, Jinghua Wang, and Yingying Zhu. Few-shot classification with contrastive learning. In *European conference on computer vision*, pages 293–309. Springer, 2022. 2, 3

[55] Jifei Yu, Xinbo Gao, Dacheng Tao, Xuelong Li, and Kaibing Zhang. A unified learning framework for single image super-resolution. *IEEE Transactions on Neural networks and Learning systems*, 25(4):780–792, 2013. 4

[56] Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In *Curves and Surfaces: 7th International Conference, Avignon, France, June 24-30, 2010, Revised Selected Papers 7*, pages 711–730. Springer, 2012. 3

[57] Kaibing Zhang, Dacheng Tao, Xinbo Gao, Xuelong Li, and Jie Li. Coarse-to-fine learning for single-image super-resolution. *IEEE transactions on neural networks and learning systems*, 28(5):1109–1122, 2016. 4

[58] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 286–301, 2018. 1

[59] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2472–2481, 2018. 1