Contrastive Learning in Aquatic Species Detection: Enhancing Aquatic Image with Super-Resolution Convolutional Neural Network and Contrastive Learning on Video Data for Classifying Invasive Aquatic Species

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

A. Problem Statement

Zebra and Quagga mussels are native to Eurasia but have become widely introduced in North American waters, causing ecological disruption [5]. These organisms fight for resources, causing the extinction of other freshwater mussels [6] and spread rapidly, forming large colonies restricting water flow and impeding power generation from water systems, clogging pipes and other machinery [1]. These invasive species can cause ecological damage by out-competing native species for food and attaching themselves to organisms, pipes, boats, and other critical infrastructure [4]. Invasive mussels adhere to hard surfaces and concrete blocks, restricting water flow through hydroelectric, irrigation, and fish facilities [2]. 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 [1, 2]. Dreissenid mussels cause hundreds of millions of dollars in damage to water infrastructure in the United States each year. They spread quickly and lay millions of eggs annually. Once established, controlling their population growth is impossible. It is crucial to monitor their presence at the larval stage. [2]. The conventional methods of detecting veliger presence are to collect plankton or water samples and then examine the selection using cross-polarized light microscopy [2] or environmental DNA [7]. Both of these methods are costly, time-consuming, and require human experts. For this reason, it is vital to develop an automated procedure to visually monitor the veliger of invasive species from water sample videos. This Project aims to classify invasive dreissenid and non-invasive larvae from videos of water samples. However, this task is challenging due to the low-resolution images captured in water sample videos, making distinguishing between invasive and non-invasive larvae difficult. This can lead to misidentification and false negatives, affecting management and control measures. To address this issue, Single Image Super Resolution (SISR) techniques can be used to enhance the resolution of these images. This research aims to validate the effectiveness of SISR techniques, particularly Super-Resolution Convolutional Neural Network (SRCNN) techniques, in the context of aquatic invasive species monitoring. The goal is to offer a novel approach to significantly improve the data quality for classifying invasive and non-invasive larvae from water samples. We will use supervised contrastive learning for classification, outperforming supervised training with cross-entropy on classification tasks. By improving the classification of these species, we can optimize management strategies to mitigate the impacts of invasive dreissenid mussels on aquatic ecosystems.

B. Methodology

B.1. proposed Method

The aim of this project is to explore two parts. Firstly, we will utilize Single-Image Super-Resolution (SISR) techniques to enhance the aquatic image dataset. Next, we will implement a contrastive learning classification task to classify invasive and non-invasive species. Finally, we will use evaluation metrics to determine whether SISR improves the accuracy of classifying aquatic species.

B.1.1 Single-Image Super-Resolution (SISR)

Single-Image Super-Resolution (SISR) upscale low-resolution images to high-resolution ones without sacrificing image quality [9]. Our methodology is based on the Super-Resolution Convolutional Neural Network (SRCNN) [8], a framework specifically designed for image super-resolution tasks.

The SRCNN model works by initially upscaling a low-resolution image to the desired size using a bicubic interpolation method. Then, the upscaled image goes through a three-layer convolutional neural network. The first layer, the patch extraction and representation layer, extracts over-

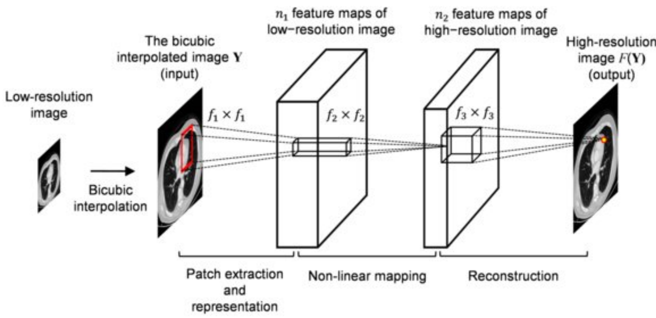

Figure 1. Architecture of the super-resolution convolutional neural network (SRCNN). Here, the architecture of SRCNN mainly consists of three components: Feature extraction, Non-linear Mapping, and Reconstruction. Here, each is responsible for extracting the features from low-resolution images, then mapping them into high-resolution features and reconstructing the final high-resolution image.

lapping patches from the upscaled image and represents each patch as high-dimensional vectors. Next, the non-linear mapping layer maps these high-dimensional vectors onto another set of vectors representing the features in the high-resolution space. Finally, the reconstruction layer synthesizes the high-resolution image from the high-dimensional vectors, enhancing the image resolution. The architecture is given in Fig1 [8]

B.1.2 Contrastive Learning Classification

Our project will employ Supervised Contrastive Learning (SCL) [3] to improve the performance of image classification models. This approach is an improvement over traditional training methods, including those that rely on cross-entropy loss. SCL works by refining how an encoder learns vector representations of input images so that images belonging to the same class are more similar to each other than those from different classes.

We will implement Supervised Contrastive Learning in two phases. In the first phase, we will train an encoder to produce vector representations of input images to maximize intra-class similarity while minimizing inter-class similarity. In the second phase, we will deploy a classifier on top of the frozen encoder for image classification.

To evaluate our approach, we will use Supervised Contrastive Learning and other deep learning-based classification tasks. This comprehensive evaluation framework will allow us to validate the superiority of SCL across a range of classification challenges, reinforcing the robustness and versatility of our approach.

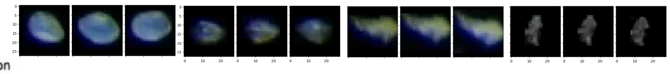

Figure 2. Example of invasive dreissenid and non-invasive species larvae in our dataset. There are four different organisms in this image, with the first six columns containing images of two different invasive larvae (three for each organism) and the following six containing images of two different non-invasive larvae.

B.2. Dataset

This study presents a specialized dataset derived from high-definition video recordings of aquatic streams. This dataset underwent meticulous processing by implementing a Kalman Filter-based, proprietary algorithm specifically designed for the identification, tracking, and subsequent extraction of larvae imagery from the source videos. We track objects in the video across frames and then extract a cropped image for each tracked object from each frame in which it appears. Each object contains images that must be classified as non-invasive or invasive. Figure 2 shows invasive and non-invasive image sets taken from a data set of water sample videos.

C. Novelty

Our study aims to explore the effectiveness of using a Super-Resolution Convolutional Neural Network (SRCNN) to upscale low-resolution images to high-resolution versions in identifying invasive and non-invasive species. Specifically, we are investigating the impact of Single-Image Super-Resolution (SISR) techniques on accurately classifying aquatic species at the larval stage.

SRCNN is chosen as our model due to its efficiency and ability to generate high-quality images without compromising on detail or introducing distortions. We want to significantly contribute to image processing by utilizing SRCNN to enhance low-resolution images to high-resolution versions.

To evaluate the performance of the SRCNN model, we will use a range of metrics, including Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Mean Squared Error (MSE). These measures will help us objectively assess the quality enhancement from low-resolution to high-resolution images, allowing us to determine the effectiveness of SRCNN in improving image fidelity.

Our approach differs from traditional training methods in that it is applied to complex encoder frameworks like ResNet and addresses multi-class classification challenges with multiple categories. We have significantly improved Super-Resolution Convolutional Learning (SCL) performance by utilizing larger batch sizes and integrating multi-layer projection heads within the encoder structure. On the other hand, for species classification, we will use the supervised contrastive method utilizing contrastive loss. Additionally, we will employ other image classification models to validate our results and evaluate our experiment extensively.

D. Timeline and Contribution

Tasks	Duration (In week)	Contribution
Project Idea Generation and project Proposal writing	1	Geethanjali and Sadia
Identifying the dataset and Preprocessing dataset	1	Sadia
Applying Image Enhancement techniques (SRCNN) and other	2-3	Geethanjali and Sadia
Observing the result and Creating new dataset with high resolution image generated by Image Enhancement techniques	2	Geethanjali
Applying the classification tasks in both datasets and Evaluating the results	2	Sadia
Applying other classification techniques	1	Geethanjali
Observing the Results	1	Geethanjali and Sadia
Final PPT making and paper writing	1	Geethanjali and Sadia

E. Expected Outcome and Worst-Case Outcome

E.1. Expected Outcome

Our research aims to improve the accuracy of aquatic species classification in their larval stages by utilizing high-resolution images produced by the Super-Resolution Convolutional Neural Network (SRCNN). The SRCNN will result in superior-quality images with enhanced detail and clarity, allowing for more precise and reliable classification of invasive and non-invasive species.

To improve the classification performance, we will adapt advanced training methodologies of contrastive learning with contrastive loss and optimize the encoder architecture with larger batch sizes and multi-layer projection heads. We aim to establish a robust framework for super-resolution imaging that supports complex classification scenarios. This could set a new benchmark for applying super-resolution techniques and pave the way for future research and applications in similar fields.

E.2. Worst-Case Outcome

In a worst-case scenario, our research might indicate that the enhancements provided by SRCNN for improving the classification accuracy of aquatic species are either minimal or insignificant enough to justify the computational resources needed for image super-resolution processing. One possible case could be that the underwater image is of poor quality and is not significantly improved after applying SRCNN, resulting in lower accuracy. Another worst-case scenario could be that the classification of invasive and non-invasive species remains the same before and after applying SRCNN. This outcome suggests that the inherent limitations in low-resolution image quality cannot adequately be compensated for by super-resolution techniques, at least insufficient to improve classification outcomes significantly. Moreover, there could be challenges in optimizing the SRCNN and SCL frameworks for our specific application, such as difficulties in tuning the model parameters to achieve the desired image quality without introducing artifacts or overfitting, which could limit the practical applicability of our approach. The performance metrics (PSNR, SSIM, MSE) may not reflect a substantial improvement in image quality or may indicate that the enhancements do not translate into better classification accuracy due to other factors, such as the complexity of the aquatic species' visual features at the larval stage.

Such findings, while disappointing, would still contribute valuable insights to the field by highlighting the limitations and challenges of applying super-resolution techniques to the classification of aquatic species. In that case, we will implement other image enhancement methods or another classification model using supervised and self-supervised methods capable of meeting such applications' stringent requirements.

References

[1] 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

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

[3] 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

[4] André L Martel and Jacqueline Baylin Madill. Twenty-six years (1990–2015) of monitoring annual recruitment of the

invasive zebra mussel (dreissena polymorpha) in the rideau river, a small river system in eastern ontario, canada. *Canadian Journal of Zoology*, 96(10):1071–1079, 2018. 1

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

[6] Don W Schloesser, Janice L Metcalfe-Smith, William P Kovalak, Gary D Longton, and Rick D Smithee. Extirpation of freshwater mussels (bivalvia: Unionidae) following the invasion of dreissenid mussels in an interconnecting river of the laurentian great lakes. *The American midland naturalist*, 155(2):307–320, 2006. 1

[7] 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

[8] Kensuke Umehara, Junko Ota, and Takayuki Ishida. Application of super-resolution convolutional neural network for enhancing image resolution in chest ct. *Journal of digital imaging*, 31:441–450, 2018. 1, 2

[9] Wenming Yang, Xuechen Zhang, Yapeng Tian, Wei Wang, Jing-Hao Xue, and Qingmin Liao. Deep learning for single image super-resolution: A brief review. *IEEE Transactions on Multimedia*, 21(12):3106–3121, 2019. 1