A Diverse Segmentation Model for Processing Fish Images in the Wild - Team 12

Qi Liu
qil059@ucsd.edu

Hengmeng Zhuang
hezhuang@ucsd.edu

Ruilong Liu
rul068@ucsd.edu

1. Abstract

The problem of accurate fish image segmentation in real environments, such as underwater conditions, is still an open issue due to noise, changes in the brightness level, and complex backgrounds [1]. Many of the current solutions are not easily transferable to unfiltered datasets, which, in turn, are usually captured in controlled environments and have uniform backgrounds [2]. To overcome these limitations, this work proposes a solution that involves the development of a U-Net-based segmentation model trained on an augmented dataset [3].

The approach consists of three stages: (1) Dataset Augmentation, in which noise, high-intensity, and low-intensity image variations are introduced; (2) Preprocessing, where Adaptive Histogram Equalization (AHE) [4], gamma correction, and Gaussian filtering [5] are applied to adjust brightness and remove noise from the images in different complex backgrounds; and (3) Model Training and Evaluation, where the U-Net model is trained on the processed dataset, and the performance of the model is evaluated.

It shows that there is a significant improvement: the model achieves 83.78% accuracy in a bright environment, compared to 81.45% in a noisy environment and 79.73% and 80.92%, respectively, when no preprocessing is applied. Additionally, the model gave an accuracy of 81.34% when tested under low-light conditions, thus proving the robustness of the model.

The proposed method enhances the segmentation performance by handling real-world issues hindering marine science and ecological monitoring progress [6].

2. Introduction

Critical marine biology tasks strongly depend on ecology and surveys as accurately as image segmentation allows, including recognition of fish species and counting and measuring biodiversity [2]. Nevertheless, providing reliable segmentation in the real environment is still challenging. It introduces many issues and challenges due to the presence of image noise, acquisition brightness variations, and complicated image backgrounds [1].

Underwater data usually need to be more organized, noisy, and consistent when compared to curated datasets. For instance, low-light conditions result from inadequate natural light penetration, especially in deep-sea environments, and overexposure due to unequal sunlight exposure in shallower waters [6]. Also, the Gaussian color noise causes image deformation [5]. Furthermore, the water's turbidity, underwater flow, and real-world camera issues affect the environment, such as underwater features, including corals, sand, or even small particles that appear in the background, making it difficult for the model to perform accurately [1].

Some of the existing segmentation models, including pre-trained networks and U-Net instances for the Fish-Vista example, have produced high-quality and well-organized results on artificial fish dataset images with transparent backgrounds, making segmentation easier [3]. However, these methods perform poorly when the data is unfiltered and more realistic, as the accuracy of the results becomes an issue [2].

These challenges highlight the weaknesses and importance of the less effective work. Effective models make ecological monitoring possible by enabling large-scale automated tools useful in marine science, conservation of biodiversity, and fisheries management [6]. For example, species identification can help detect changes in ecology due to global warming, overfishing, or pollution [1]. Furthermore, automated segmentation saves significant amounts of time and effort. Such solutions can also be applied to Autonomous Underwater Vehicles (AUVs) for image analysis and segmentation, developed by marine biologists and conservationists [6].

This work proposes a solution that overcomes these limitations, providing a real-time analysis model to improve the segmentation of fish images in dynamic and challenging underwater environments. The proposed model is capable of handling noisy data, low-light conditions, and bright underwater surveillance images. With this, the solution enhances ecological analysis systems and contributes to the conservation of marine environments under both artificial and natural conditions.

3. Previous Methods

Accurate fish image segmentation in underwater conditions has long been a problem. Several strategies have been tested, from basic image segmentation approaches to more advanced pipeline-based preprocessing systems. In this section, we detail two significant approaches: SVM-based segmentation and the Fish-Vista pipeline.

3.1. SVM-Based Segmentation

Support Vector Machines have been extensively used as a traditional approach for image segmentation, especially for simple segmentation tasks. SVM works based on featuring space and does well in simple segmentation tasks with distinct object boundaries [7].

Pros: SVM is faster and less complicated to use for small data sets. It produces the best results when clear boundaries of the objects are visible. It consumes fewer resources than deep learning models, which makes it appropriate for basic tasks.

Cons: However, SVM is not capable of dealing with noisy, high-dimensional real-world image environments, such as underwater scenarios. Real-life fish images can be affected by lighting changes, may contain noise, and have complicated backgrounds, which reduce the effectiveness of SVM. Further, SVM does not scale well when applied to large, high-resolution datasets.

3.2. Fish-Vista Pipeline

The Fish-Vista pipeline is the current data processing approach developed to work with big data sets of fish images. The Fish-Vista prepares the raw images to prepare several stages. The used pipeline consists of machine learning several tasks. Stages The pipeline follows:

1. **Deduplication**: Duplicated images are identified and removed using the MD5 checksum matching to eliminate the duplicates.
2. **Metadata Filtering**: Quality metadata, such as fish visibility and orientation, is used to filter out irrelevant or incomplete images.
3. **Species Name Resolution**: Open Tree Taxonomy cleans up species names from noisy or placeholder values, resulting in correct and well-formatted names.
4. **Cropping and Background Removal**: State-of-the-art models for cropping fish specimens and eliminating complicated backgrounds include Grounding DINO and Segment Anything Model (SAM) [2].

Pros: Fish-Vista ensures that only high-quality datasets are used for training by providing clean, suitably annotated, and filtered images. It is very useful in filtering elements such as tags, rulers, or other noise that can interfere with the segmentation process.

Cons: Nevertheless, Fish-Vista has a significant drawback in that it depends on data preprocessing to enhance the quality of the datasets. This method does not address the issue of identifying the regions of interest in the raw non-processed images; therefore, the effectiveness of the method is compromised in practical applications where there is a lot of interference and variability in the data.

3.3. Conclusion of Limitation

SVM-based approaches can be considered a traditional baseline that is not effective for challenging conditions in the underwater environment. On the other hand, Fish-Vista is most effective in data preparation since it helps create clean and properly annotated datasets. However, this does not solve the problem of how to segment noisy and unfiltered images which is an issue when working with real-life underwater data.

4. Proposed Method

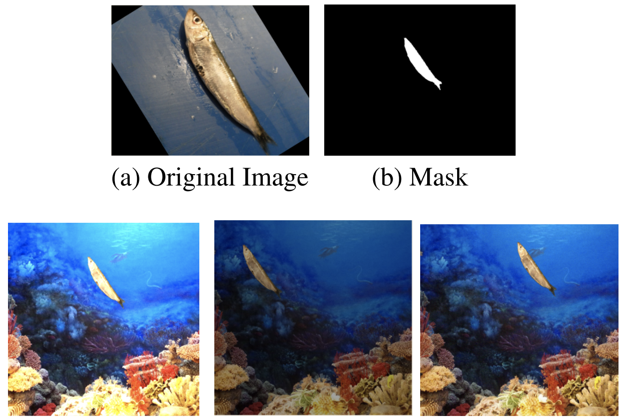
Our proposed method consists of two key components: dataset preprocessing and model development.

4.1. Data Preprocessing

In order to simulate the real world and create a robust segementation model, we enhance the Fish-Vista dataset by introducing diverse backgrounds and adjust the intensity and noise injection[2]. Precisely, as shown in Figure2, the steps are as follows:

1. **Fish Isolation:** Extract fish images from the original clear and clean library dataset using segmentation masks to produce isolated fish images.
2. **Background Build Up:** Collect background images that simulate real world environment including oceanic, rocky, and colorful coral reef scenes. Combine the isolated fish images with the modified background images to create a new dataset. The fish is put in a random place with a random size after rotations.
3. **Background Modification:** After combining the pictures, apply two modifications to enhance our dataset:
   - **Brightness Adjustments:** Apply brightness changes to form dataset with low-light and overexposed backgrounds.This simulate the unedited undersea photos.
   - **Noise Injection:** Add Gaussian noise to simulate the photo taken underwater and having environmental disturbances.

This preprocessed dataset consists of a wide range of lighting conditions and noise levels which help to improve the model's adaptability to



(a) Original Image (b) Mask

(c) Exposed (d) Underexposed (e) noise

Figure 1. Processed Example

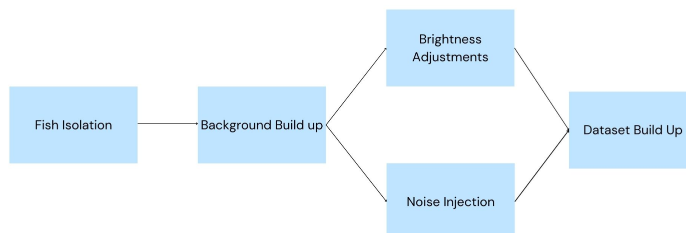


Figure 2. Dataset Build Up

complex real-world scenarios.

4.2. Model Development

As shown in Figure3, our segmentation model contains two stage, preprocessing stage and U-Net deep learning model stage:

1. **Preprocessing Stage:**
   - **Intensity Adjustments:** First, our model will assess the average intensity of the image after changing it to grey scale. If the intensity is too high, our model increase the clip limit parameters and decrease the gamma value to balance the brightness. On contrast, if the intensity is too low, our model lower the clip limit parameters and increase the gamma value to enhance visibility. By applying the gamma transformation and Adaptive Histogram Equalization techniques, our model improve the visibility of fish against complex backgrounds and also deal with the underexposed or overexposed fish image.
   - **Denoising:** To reduce the potential noise in the images, our model use Gaussian and median filtering. The denoising process includes noise detection and filter application. Our model first apply a smooth filter (Gaussian filter) to the image and analyze the difference between the original image and the filtered image. If the difference shows a significant deviation, Gaussian noise is existed. If Gaussian noise is detected, our model uses Gaussian filters to reduce the noise that can improve the performance of the segmentation model.
2. **Deep Learning Model:** Our model then applies the state-of-the-art segmentation network U-Net on the preprocessed dataset[3].

5. Experiment and Result

In this section, I describe in detail the results achieved by the Fish-Segmentation Model (FishSeg Model). In the Dataset Augmentation stage, in order to simulate the living environment of fish in nature, FishSeg Model will not only use a complex ocean background map, but also randomly increase the Gaussian noise and increase the brightness of all images to increase the complexity of the images. Therefore, FishSeg Model will first perform a series of image preprocessings on the data images to reduce the influence of noise or intensity on image segmentation. Below are the results of FishSeg Model's data preprocessing of the dataset using image processing.

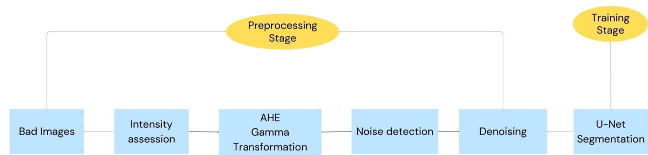
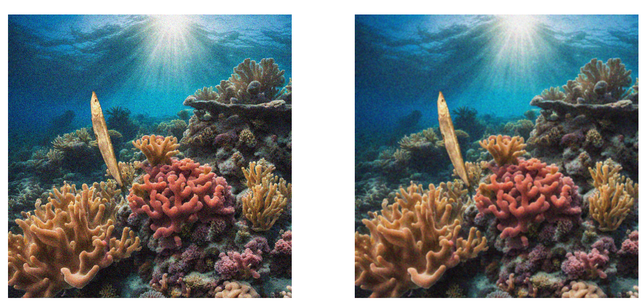


Figure 3. Image Processing Stages

As shown in Figure 4, these two graphs represent the comparison images before and after data preprocessing, respectively. Figure 4a represents the dataset with Gaussian noise, and Figure 4b shows that FishSeg Model will first detect whether there is Gaussian noise in the input image, and if so, will perform Gaussian denoising on the input image.

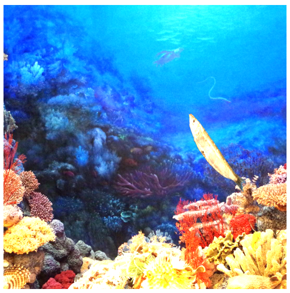


(a) Image before Gaussian Denoising

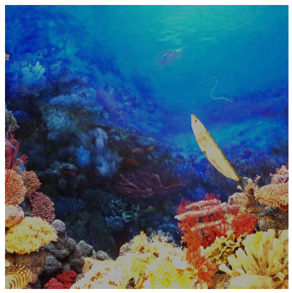
(b) Image after Gaussian Denoising

Figure 4. Data with Gaussian Noise

As shown in Figure 5, the two graphs represent a comparison of the treatment of overbrightness. Figure 5a represents the dataset with excessive brightness, and Figure 5b represents FishSeg Model, which will first detect whether the input image will be too bright, and if so, will adjust the brightness of the input image.



(a) Image before Brightness Adjustment

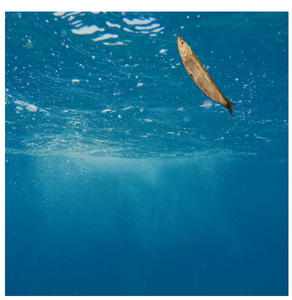


(b) Image after Brightness Adjustment

Figure 5. Data with Excessive Brightness

As shown in Figure 7, the two graphs represent a comparison of the treatment of overdarkness. Figure 6a represents the dataset with excessive darkness, and Figure 6b represents FishSeg Model, which will first detect whether the input image will be too dark, and if so, will increase the intensity of the input image.

(a) Image before Darkness Adjustment

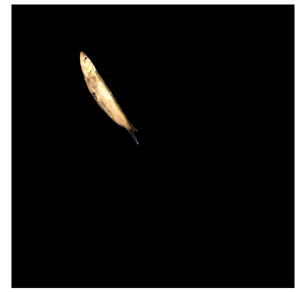


(b) Image after Darkness Adjustment

Figure 6. Data with Excessive Darkness

The above is the result of FishSeg Model's preprocessing of model data, and then I will show the training results and accuracy of FishSeg Model. First, Figure 7a is the raw data before being entered into the FishSeg Model, and Figure 7b is a picture of the individual fish that were segmented by the FishSeg Model. As we can roughly see from the results, the FishSeg Model has a good segmentation performance and correctly separates the fish from the complex background image.

(a) Image before Segmentation



(b) Image after Segmentation

Figure 7. Segmentation Result

Figure 8 shows the accuracy of each epoch of FishSeg Model during training. As we can see, the accuracy of FishSeg Model increased steadily during training, but eventually converged between 80% and 85%, reaching a decent accuracy of 83%.

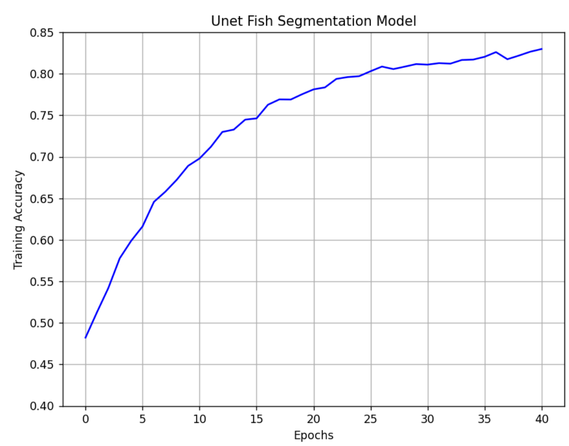


Figure 8. Accuracy of FishSeg model

Not only does the FishSeg Model itself achieve a commendable accuracy, but we also compared the FishSeg Model with normal segmentation models and found that the FishSeg Model has higher accuracy than the other models. As shown in Table 1, the first row represents the ordinary segmentation model without digital image preprocessing of the data, and its accuracy is 79.73% and 80.92% in the dataset with Gaussian noise and excessive brightness, respectively. The accuracy of FishSeg Model in the dataset with Gaussian noise and excessive brightness is 81.45% and 83.78%, respectively. Although this improvement is not obvious, we can still see that pre-digital image processing of data can improve the accuracy of image segmentation in complex environments.

Model	Noise Accuracy	Bright Accuracy
Without Processing	79.73	80.92
With Processing	**81.45**	**83.78**

Table 1. Accuracy of Different Models

Table 2 illustrates the robustness of FishSeg Model in an environment that has never been seen before. Since the FishSeg Model was only trained on datasets with Gaussian noise and overly bright conditions, but not on overly dark conditions, I added overly dark conditions to the test set to explore whether the FishSeg Model has the ability to handle completely new situations. The first row of Table 2 shows the accuracy of the FishSeg Model in the overly bright test set, and the second row shows the accuracy of the FishSeg Model in the overly dark test set. It can be observed that, although the accuracy in the second row is not particularly high, it still achieves a reasonably good result. It illustrates the robustness of FishSeg Model in different environments.

Testing Condition	Accuracy
Bright Condition	83.78
Dark Condition	81.34

Table 2. Accuracy of Test Condition

6. Discussion

In summary, we have successfully developed an augmented dataset by generating diverse backgrounds and add various level of noise and intensity to it. This enhanced dataset is more close to the real world environment and help to develop a more robust segmentation model. The model has achieved good accuracy improvement in different conditions. Even using only under-exposed image to test the overexposed-image-trained model, it still shows a good improvement.

However, there are still some limitations in our model. The accuracy improvements are significant as expected. TWe find that our model do not perform well when dealing with small fish detection. Additionally, the model may not fully simulate some of the real-world underwater environments, such as motion blur and different water clarity.

6.1. Pros of our method:

- **Improved Robustness:** Our dataset has complex backgrounds, brightness levels, and noise level. It makes the model more robust when facing the real-world environment.
- **Exposed Handling:** The Adaptive Histogram Equlization and gamma transformation significantly improves the contrast and visibility of our target fish in low-intensity and noisy conditions.
- **Noise Handling:** The combination of Gaussian and median filtering helps reduce different level of noise.

Cons:

- **Limited Performance on Small Fish Detection:** The model still does not have a good performance with detecting smaller fish. It might because it lacks small-scale features in the training data.
- **Limited Performance in Complex Environments:** The model does not fully simulate the complexities in underwater environments, such as motion blur and different water conditions. This could affect the accuracy when using the

data from a real world scenarios.

- **Slight Improvement:** While the model shows some improvement, the accuracy gains are not as good as expected.

6.2. Possible Improvements for Future Works

- **Edge-Preserving Filters:** Future work can do more on edge-preserving filters. These filters can help preserve object boundaries while reducing noise. It might improve the segmentation model more.
- **Motion Blur Handling:** The images underwater may includes a lot of motion blur. Future works can focus more on motion deblurring algorithms. It could improve the model's performance in dynamic underwater environments.
- **Data Augmentation for Small Objects:** Future works can focus more on small fish or objects. Capture more small features. can help improve detection accuracy for these smaller targets.

7. Conclusion

In order to better split the image of the fish from the background image alone, we create FishSeg Model. FishSeg Model has the ability to face complex images, not only in the case of varying degrees of Gaussian noise in the background image but also in the case of overly bright images. Compared with other different models, FishSeg Model also has certain advantages. In addition, the FishSeg Model has the ability to deal with new environments, and it still maintains a good accuracy rate when tested in a dataset that has never been seen before. It can be seen that FishSeg Model has the ability to deal with different environments.

Contribution: Qi Liu : Hengmeng Zhuang : Ruilong Liu = 40% : 30% : 30%

References

[1] H. Jiang, X. Liu, and W. Zhou, "Revolutionizing underwater imaging: Unveiling the impact and challenges of image segmentation," *Springer Underwater Image Analysis Journal*, 2023. 1, 2

[2] K. S. Mehrab, M. Maruf, and A. e. a. Daw, "Fish-vista: A multi-purpose dataset for understanding & identification of traits from images," *arXiv preprint arXiv:2407.08027*, 2024. 1, 2, 3

[3] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," pp. 234–241, 2015. 1, 2, 4

[4] S. M. Pizer, J. S. Amburn, and J. D. e. a. Austin, "Adaptive histogram equalization and its variations," *Computer Vision, Graphics, and Image Processing*, vol. 39, no. 3, pp. 355–368, 1987. 1

[5] Y. Zhang and Z. Zhang, "An adaptive gaussian filter for noise reduction and edge detection," *IEEE Transactions on Image Processing*, vol. 17, no. 7, pp. 1186–1195, 2008. 1

[6] Z. Lian, Y. Zhang, and S. Zhang, "Watermask: Instance segmentation for underwater imagery," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1458–1467, 2023. 1, 2

[7] O. Ulucan, D. Karakaya, and M. Turkan, "A large-scale dataset for fish segmentation and classification," in *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*, pp. 1–5, IEEE, 2020. 2