Just Scratching the Surface: Exploring Underwater Image Enhancement with CLIP

Jingwen Wang, Zixuan Hu
Department of Computer Science, University of Toronto

Motivation

- Improving underwater images is essential for marine research and exploration.
- Conventional techniques like **WaterNet[1]** depend on raw-reference image mapping, where the enhancement quality is limited by reference images.
- With fixed references, there is a significant challenge in modifying enhancements to align with specific or new requirements.
- Recent advancements with the **CLIP[2]** model showcase its ability to bridge images and text, understanding and correlating visual content with linguistic descriptions.

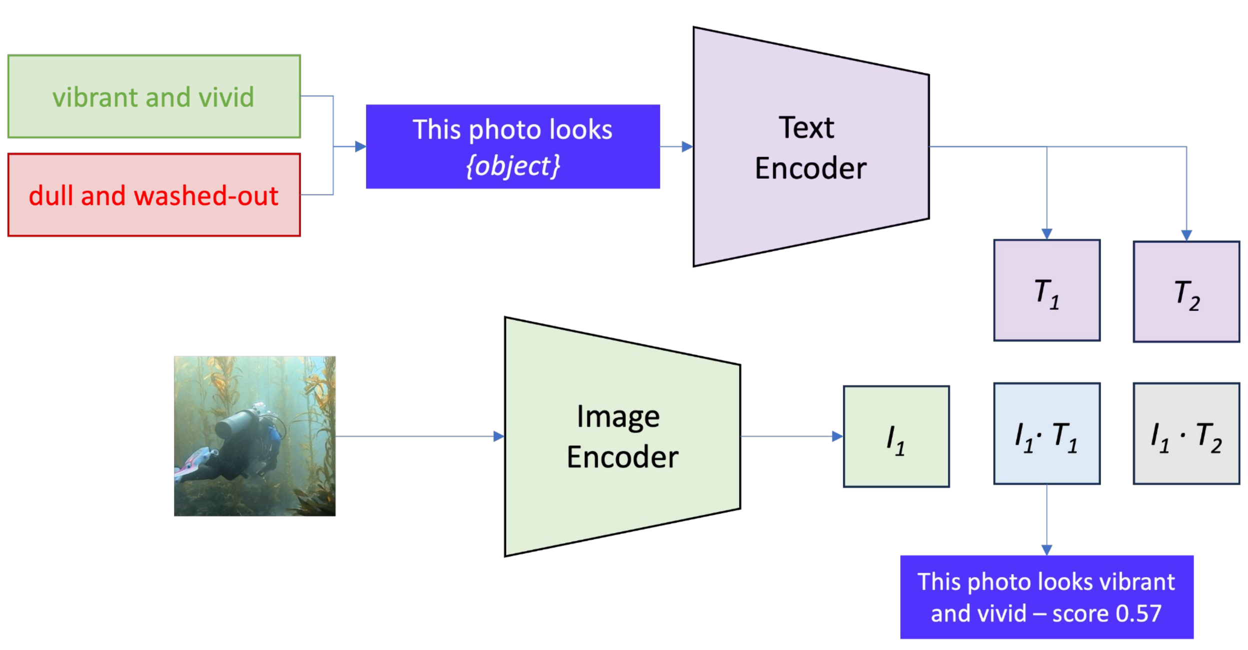

Figure 1. CLIP model: Assessment of image quality by contrasting pairs.

- The pretrained CLIP model can effectively assess image quality, distinguishing between high-quality and low-quality images, due to its training on extensive and diverse datasets [3].
- We intend to enhance WaterNet by incorporating CLIP, aiming to improve image colors, precisely adjust white balance, and evenly balance exposure, solely using the **UIEB dataset[1]**.

Related Work

- WaterNet[1] was proposed alongside the **UIEB** dataset. It excels in reconstructing underwater images to closely match reference images.
- The CLIP[2] model represents a breakthrough in contextual understanding of images. Trained on extensive image-text pairs, it has the unique ability to correlate visual content with textual descriptions.
- The CLIP-LIT[3] underscores CLIP's potential in unsupervised backlit image enhancement.
- LSUI Dataset[4]: Encompassing 4,279 paired underwater images, LSUI offers a comprehensive range of scenes, lighting conditions, water types, and target categories.

References

[1] Li et al., *An underwater image enhancement benchmark dataset and beyond.* IEEE Transactions on Image Processing, 29:4376– 4389, 2019.
[2] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. *Learning transferable visual models from natural language supervision.* In International conference on machine learning, pages 8748–8763. PMLR, 2021
[3] Zhexin Liang, Chongyi Li, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. *Iterative prompt learning for unsupervised backlit image enhancement.* In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 8094–8103, 2023.
[4] Lintao Peng, Chunli Zhu, and Liheng Bian. *U-shape transformer for underwater image enhancement.* IEEE Transactions on Image Processing, 2023.

Methods

Prompt design:

1. **Colorfulness: "***Vibrant and Vivid",* to restore the natural color often dimmed by water.
2. **White Balance:** *"Accurate Color representation",* to correct the prevalent blue/green color cast.
3. **Exposure:** *"Well-lit and clear",* to discern fine details and ensure the scene's clarity.

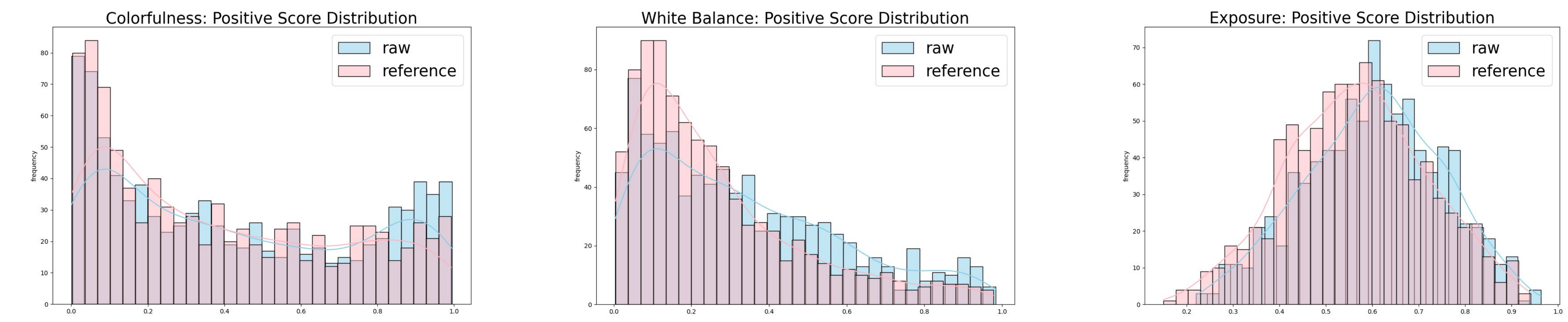

Figure 2. the distribution of positive scores for contrastive pairs, comparing raw and reference images from the **UIEB** dataset.

Overview of the Mode:

Our approach employs a pretrained WaterNet as the base model, subsequently enhanced through a dual loss strategy with CLIP's contrastive pair and perceptual losses. The model is trained on the **UIEB** dataset with various contrastive pairs and then evaluated using the **LSUI** dataset [4].

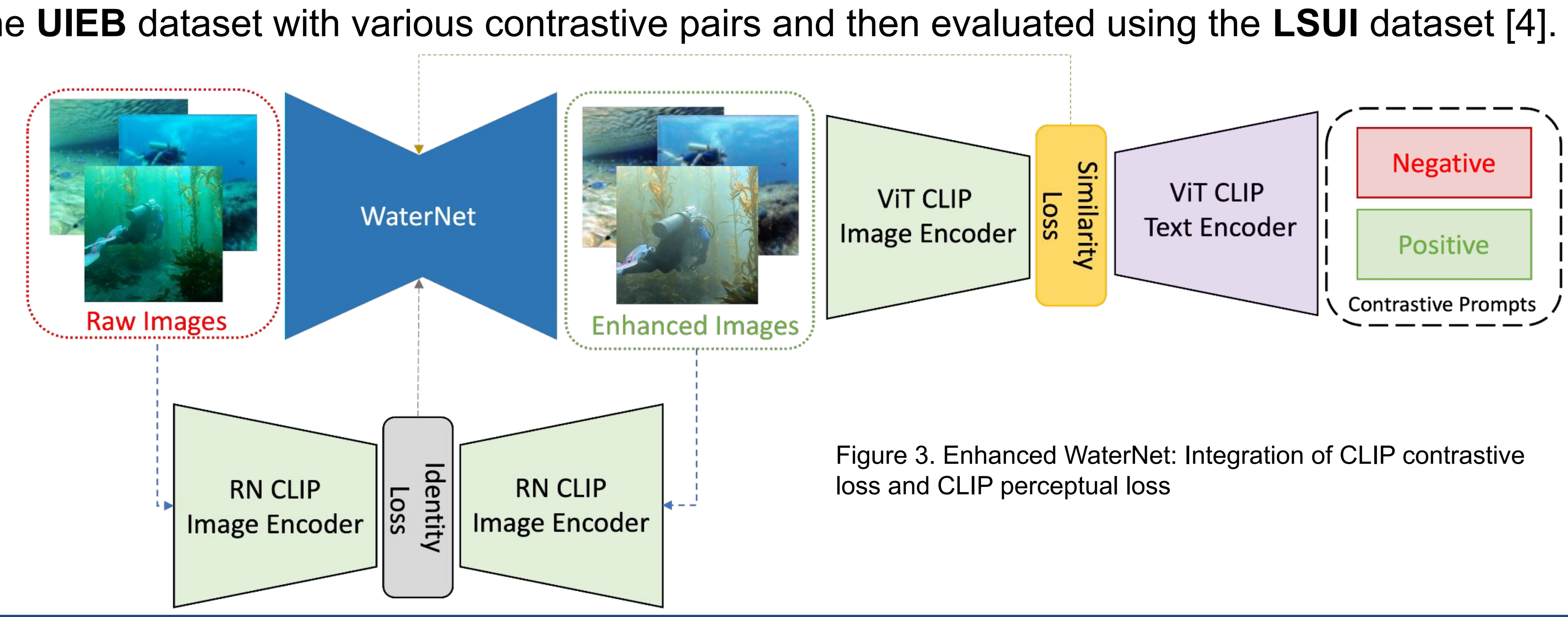

Figure 3. Enhanced WaterNet: Integration of CLIP contrastive loss and CLIP perceptual loss

Experimental Results

model	psnr	ssim	colorfulness	white balance	exposure
base	**22.154**	**0.850**	0.489	0.354	0.548
Color-Enhanced	20.421	0.822	**0.682**	0.310	0.558
WhiteBalance-Enhanced	20.490	0.839	0.332	**0.549**	0.559
Exposure-Enhanced	21.202	0.830	0.407	0.230	**0.618**

Table 1. the performance metrics of various models on the **LSUI dataset.** It uses PSNR and SSIM for comparative analysis with reference images. It also includes positive scores for colorfulness, white balance, and exposure, based on contrastive pairs of images produced by each model.

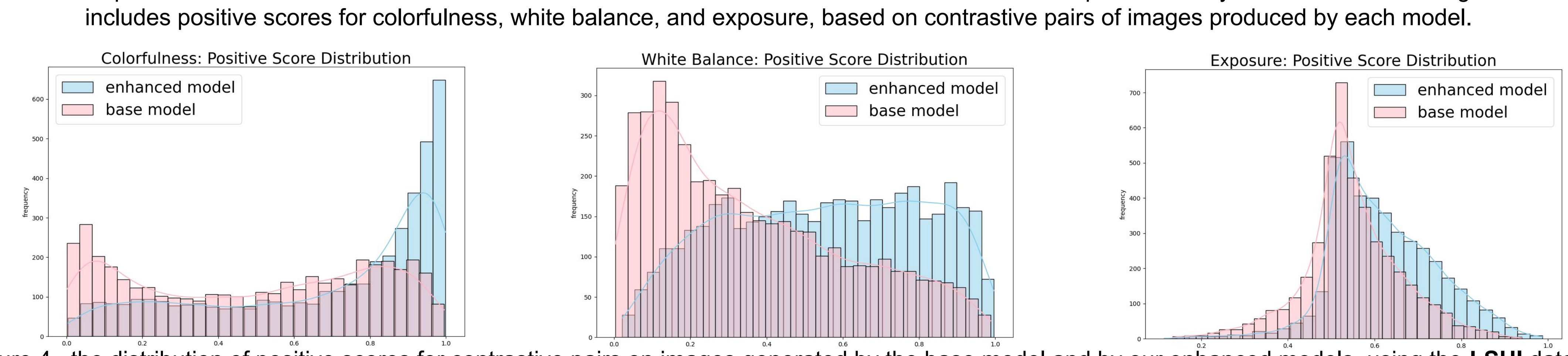

Figure 4. the distribution of positive scores for contrastive pairs on images generated by the base model and by our enhanced models, using the **LSUI** dataset.

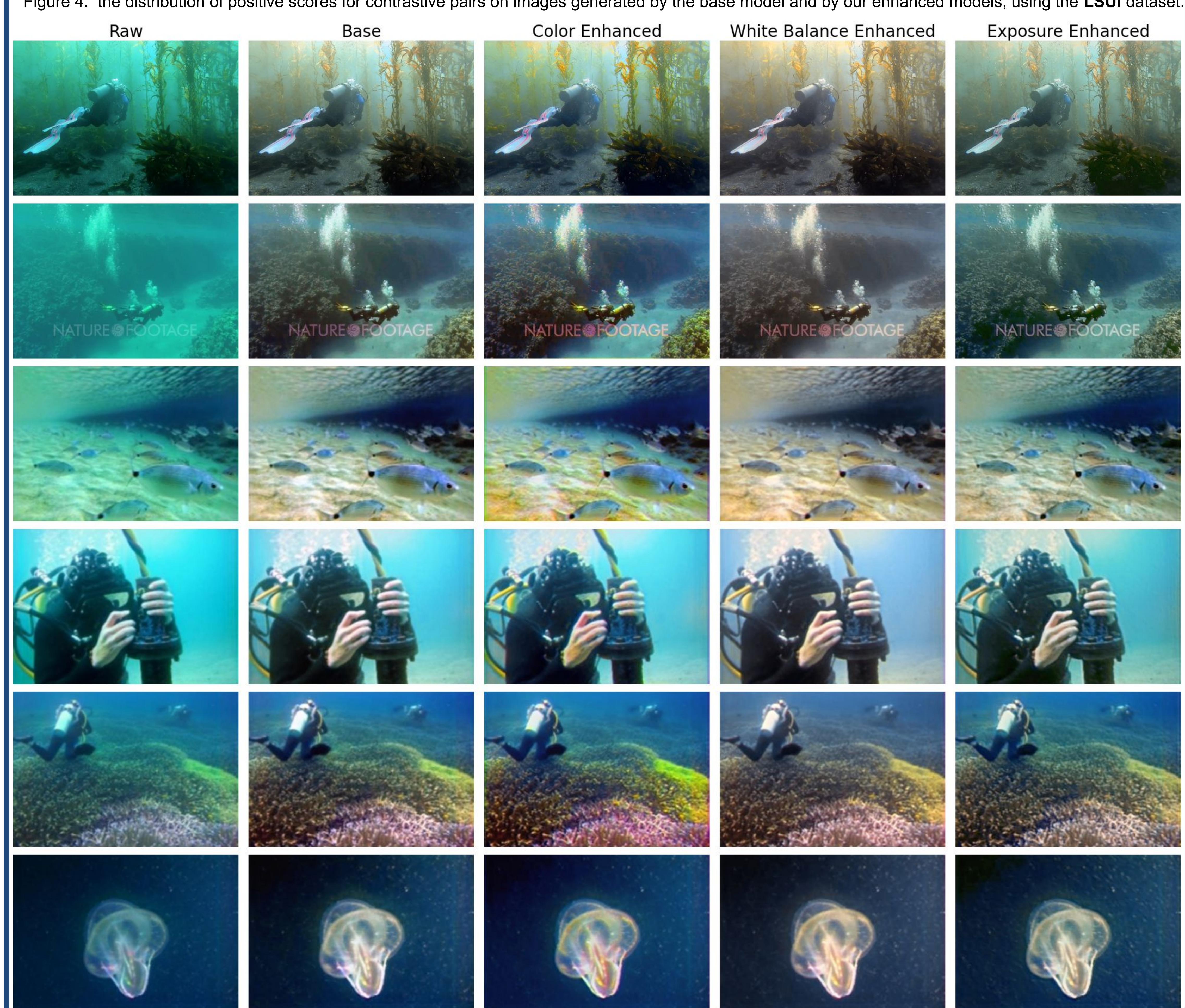

Figure 5. Enhanced underwater images. Color, white balance, and exposure enhancement are processed by different enhanced waternet.