Supplementary Material: AutoEnhancer: Transformer on U-Net Architecture search for Underwater Image Enhancement

Yi Tang[1], Takafumi Iwaguchi[1], Hiroshi Kawasaki[1], Ryusuke Sagawa[2], and Ryo Furukawa[3]

[1] Kyushu University
tang.yi.727@m.kyushu-u.ac.jp, {iwaguchi,kawasaki}@ait.kyushu-u.ac.jp
[2] National Institute of Advanced Industrial Science and Technology
ryusuke.sagawa@aist.go.jp
[3] Kindai University
furukawa@hiro.kindai.ac.jp

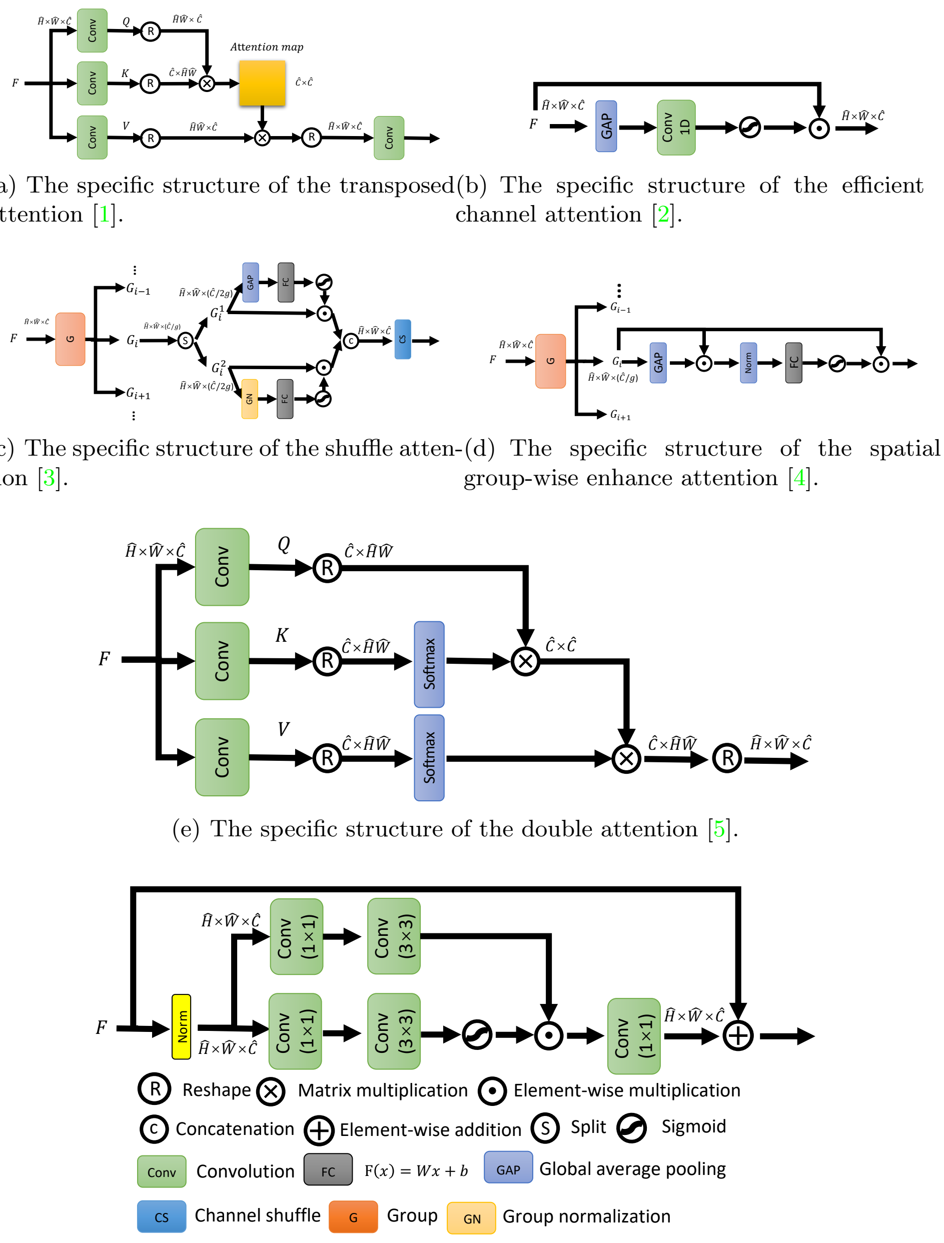
1 Implementation

In this section, we report a further description of our implementation of our NAS-based transformer. By replacing the multi-head attention module, we can derive five different transformer structures: transformer with transposed attention [1] (T_{ta}), transformer with efficient channel attention [2] (T_{eca}), transformer with shuffle attention [3] (T_{sa}), transformer with spatial group-wise enhance attention [4] (T_{sge}) and transformer with double attention [5] (T_{da}). Here, we present their specific attention structure in Figure. 1. For more details, please read their papers. Through these modules, the input with arbitrary resolutions can be directly fed into the corresponding transformer and their scale of parameters is small as well.

2 Extra experiments

We also evaluate our methods by different training setting: Ushape setting [6] and RCTNet setting [7]. The biggest difference between these settings is the training data. As for the Ushape setting, the networks are trained only on the training set from the LSUI dataset. The RCTNet setting is to train the networks on different datasets. That is to separately train the networks with the respective training data from the datasets. As shown in Table. 1, our deep model can achieve remarkable performance in different settings. Moreover, the performance with Ushape setting is better. On one hand, it denotes that the LSUI dataset is indeed a high-quality dataset, which can provide abundant scenes and diverse objects for the training of data-driven models. On the other hand, the proposed network is able to make the most of these data, thus extracting robust and reliable features for enhancement.

We also present more visual results in Figure. 2. Moreover, we present some failure cases, which cannot completely recover the original color or content by



(a) The specific structure of the transposed attention [1].

(b) The specific structure of the efficient channel attention [2].

(c) The specific structure of the shuffle attention [3].

(d) The specific structure of the spatial group-wise enhance attention [4].

(e) The specific structure of the double attention [5].

(f) The specific structure of the deployed feed-forward function.

Fig. 1: The specific structures of different self-attention modules and feed-forward function.

Table 1: Quantitative results by using different training settings.

Training setting	UIEB		LSUI		EUVP	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Ushape setting [6]	25.45	0.9231	26.13	0.8608	29.56	0.8818
RCTNet setting [7]	22.82	0.9137	-	-	26.59	0.8451

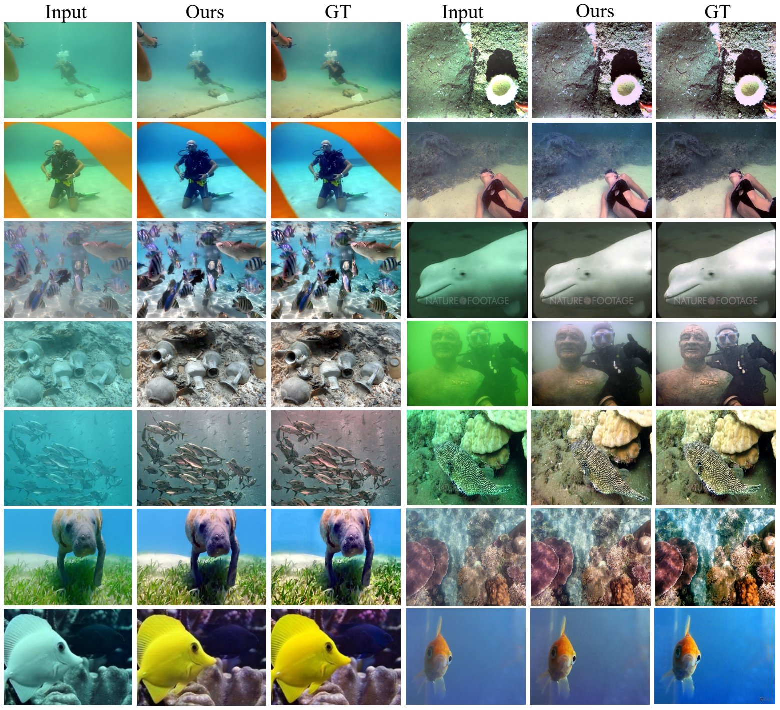


Fig. 2: More Visual results on testing datasets.

the proposed approach. For example, in the fifth row of Figure. 2, we can see the ground truth presents a red color style. Our enhanced images can recover part of them but fail in the entire image. For these colorful marine lives, there are few samples in the datasets. It is still difficult to capture the diverse color information by using few training data.

In the Table. 2, we also report the runtime comparison. Among the recent deep learning-based methods, the model size and runtime are competitive.

3 Application on underwater detection

In order to validate the supportive function of our enhancer, we use a detector [13] to detect the underwater objects by using an original video and the corresponding enhanced one. The qualitative results are shown in Figure. 3. As we

Table 2: Runtime and model sizes of the deep learning-based methods

Method	WaterNet [8]	FUnIE [9]	UGAN [10]	UIE-DAL [11]	Ucolor [12]	Ushape [6]	Ours
Param.	25M	7M	57M	19M	157M	66M	12M
Time	0.55s	0.02s	0.06s	0.04s	1.87s	0.04s	0.02s

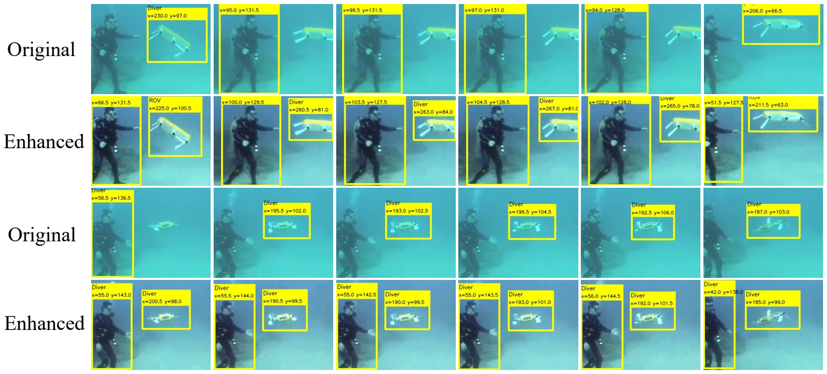


Fig. 3: The supportive function of the the proposed enhancer for object detection in underwater scenarios.

can see, the visual results of using the enhanced video are better than the original inputs. As for the multiply objects in the video, not all of the objects can be highlighted by the detector by using the original video. After the enhancement by the proposed approach, the video frames turn clearer, which is very useful to help the detector extract high-level features, thus generating accurate bounding boxes of the objects. Moreover, our runtime is 0.02s per frame. It is a small burden for the detector.

References

1. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H.: Restormer: Efficient transformer for high-resolution image restoration. arXiv preprint arXiv:2111.09881 (2021)
2. Qilong Wang, Banggu Wu, P.Z.P.L.W.Z., Hu, Q.: Eca-net: Efficient channel attention for deep convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2020)
3. Zhang, Q.L., Yang, Y.B.: Sa-net: Shuffle attention for deep convolutional neural networks. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE (2021) 2235–2239
4. Li, X., Hu, X., Yang, J.: Spatial group-wise enhance: Improving semantic feature learning in convolutional networks. arXiv preprint arXiv:1905.09646 (2019)
5. Chen, Y., Kalantidis, Y., Li, J., Yan, S., Feng, J.: Aˆ 2-nets: Double attention networks. Advances in neural information processing systems **31** (2018)
6. Peng, L., Zhu, C., Bian, L.: U-shape transformer for underwater image enhancement. arXiv preprint arXiv:2111.11843 (2021)

7. Kim, H., Choi, S.M., Kim, C.S., Koh, Y.J.: Representative color transform for image enhancement. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. (2021) 4459–4468
8. Li, C., Guo, C., Ren, W., Cong, R., Hou, J., Kwong, S., Tao, D.: An underwater image enhancement benchmark dataset and beyond. IEEE Transactions on Image Processing **29** (2019) 4376–4389
9. Islam, M.J., Xia, Y., Sattar, J.: Fast underwater image enhancement for improved visual perception. IEEE Robotics and Automation Letters **5** (2020) 3227–3234
10. Fabbri, C., Islam, M.J., Sattar, J.: Enhancing underwater imagery using generative adversarial networks. In: Proceedings of the IEEE International Conference on Robotics and Automation, IEEE (2018) 7159–7165
11. Uplavikar, P.M., Wu, Z., Wang, Z.: All-in-one underwater image enhancement using domain-adversarial learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. (2019) 1–8
12. Li, C., Anwar, S., Hou, J., Cong, R., Guo, C., Ren, W.: Underwater image enhancement via medium transmission-guided multi-color space embedding. IEEE Transactions on Image Processing **30** (2021) 4985–5000
13. Islam, M.J., Fulton, M., Sattar, J.: Toward a Generic Diver-Following Algorithm: Balancing Robustness and Efficiency in Deep Visual Detection. IEEE Robotics and Automation Letters (RA-L) **4** (2018) 113–120