Journal of computer science and technology: Instruction for authors. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY

Supplementary Materials

1 Salient Object Detection

We employ MSRA-10K [3] as the training set, and take MSRA-B [1], DUTs [2], and DUT-OMRON [2] as the validation set. All the images in these datasets are resized to 224×224 for the training and test. Three models including UNet [4], LinkNet [5], and FPN [6] are involved in the evaluation. We take ResNet50 [7] as the backbone network for these models and the implementation details can be found here[1]. Besides, we take the intersection over union (IoU), F_1-measure (F_1), E-measure (EM) [8], and mean absolute error (MAE) as the evaluation metrics. The implementations of IoU, F_1 and MAE can be found here[1], and EM refers to [8]. We employ ClipQ for model quantization and do not quantize the first and last layers of models. We set the average bitwidth as 2 for QPL to apply an extremely low-bitwidth quantization for these large SOD models for great computing benefits.

The quantitative comparison results are in Table 1. The model size and memory footprint of UNet are reduced by 14.88× and 172.56×, respectively, when it is quantized into the average bitwidths of 2.15/2.76 with AutoQNN, and the performance degradation in four metrics across all datasets is less than 0.01. AutoQNN also shrinks the memory footprint of FPN and LinkNet by 155.67× and 172.64× through quantizing models from full precision into low precision with average bitwidths of 2.20/2.99 and 2.20/2.65, respectively, and only incurs slight performance degradation which is not more than 0.018 in all metrics across three datasets. The results demonstrate that QPL can reduce memory footprint by using the extremely low-bitwidth weight and activation with negligible performance degradation. This highlights that QPL can search for a suitable quantizing strategy to maintain comparable feature extraction quality of SOD models.

[1] https://github.com/qubvel/segmentation_models, Oct. 2022.

References

[1] Liu T, Yuan Z, Sun J, Wang J, Zheng N, Tang X, Shum H. Learning to detect a salient object. *IEEE TPAMI*, 2011, 33(2):353–367. DOI: 10.1109/TPAMI.2010.70.

[2] Wang L, Lu H, Wang Y, Feng M, Wang D, Yin B, Ruan X. Learning to detect salient objects with image-level supervision. In *IEEE CVPR*, 2017, pp. 136–145. DOI: 10.1109/cvpr.2017.404.

[3] Cheng M M, Mitra N J, Huang X, Torr P H, Hu S M. Global contrast based salient region detection. *IEEE TPAMI*, 2014, 37(3):569–582. DOI: 10.1109/cvpr.2011.5995344.

[4] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, volume 9351 of *Lecture Notes in Computer Science*, 2015, pp. 234–241. DOI: 10.1007/978-3-319-24574-4_28.

[5] Chaurasia A, Culurciello E. Linknet: Exploiting encoder representations for efficient semantic segmentation. In *VCIP*, 2017, pp. 1–4. DOI: 10.1109/vcip.2017.8305148.

[6] Lin T, Dollár P, Girshick R B, He K, Hariharan B, Belongie S J. Feature pyramid networks for object detection. In *IEEE CVPR*, 2017, pp. 936–944. DOI: 10.1109/cvpr.2017.106.

Table 1. Evaluation Results of the Salient Object Detection Models on Four Metrics

Models	W/A	MSRA-B [1]				DUTs [2]				DUT-OMRON [2]			
		IoU↑	F_1↑	EM↑	MAE↓	IoU↑	F_1↑	EM↑	MAE↓	IoU↑	F_1↑	EM↑	MAE↓
UNet	32/32	0.8936	0.9437	0.9645	0.0244	0.6106	0.7575	0.8417	0.0773	0.6390	0.7779	0.8262	0.0696
	2.15/2.76	0.8850	0.9389	0.9616	0.0265	0.6074	0.7549	0.8358	0.0775	0.6390	0.7779	0.8236	0.0688
FPN	32/32	0.8920	0.9428	0.9644	0.0248	0.6286	0.7713	0.8505	0.0737	0.6316	0.7729	0.8292	0.0732
	2.20/2.99	0.8908	0.9422	0.9644	0.0251	0.6298	0.7720	0.8469	0.0719	0.6495	0.7865	0.8346	0.0662
LinkNet	32/32	0.8904	0.9419	0.9640	0.0252	0.6286	0.7713	0.8519	0.0725	0.6426	0.7809	0.8386	0.0690
	2.20/2.65	0.8858	0.9393	0.9619	0.0262	0.6234	0.7673	0.8425	0.0726	0.6438	0.7821	0.8272	0.0673

Note: The ↑ denotes that the higher values shows better results and the ↓ vise versa. W/A idicates the average bitwidth for weights and activation, respectively.

[7] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In *IEEE CVPR*, 2016, pp. 770–778. DOI: 10.1109/cvpr.2016.90.

[8] Fan D P, Gong C, Cao Y, Ren B, Cheng M M, Borji A. Enhanced-alignment measure for binary foreground map evaluation. In *IJCAI*, 2018, pp. 698–704. DOI: 10.24963/ijcai.2018/97.