Multi-Scale Denoising in the Feature Space for Low-Light Instance Segmentation

Joanne Lin [1] Nantheera Anantrasirichai [1] David Bull [1]

[1]Visual Information Laboratory, University of Bristol

Introduction

Low-light images are heavily affected by noise from a low photon count in darker conditions, making the task much more challenging. Common approaches include applying a pre-processing step first to enhance the image, before passing into existing methods.

We propose our "plug-and-play" weighted non-local blocks (wNLB) into backbones of architectures for an end-to-end low-light instance segmentation method.

Proposed Method

We build upon the existing non-local blocks (NLB) [1] by adding a learnable parameter w. This allows the network to control the level of feature denoising at different scales, as seen in Fig. 1.

Our wNLBs computes the following:

$$\mathbf{z} = wW_z\mathbf{y} + (1-w)\mathbf{x} \quad (1)$$

where $\mathbf{x}$ is the input feature map, $\mathbf{y}$ is the output from the NL means operation, W_z is the weight matrix from the 1×1 convolutional layer after the NL means operation and $\mathbf{z}$ is the output. This is shown in Fig. 2.

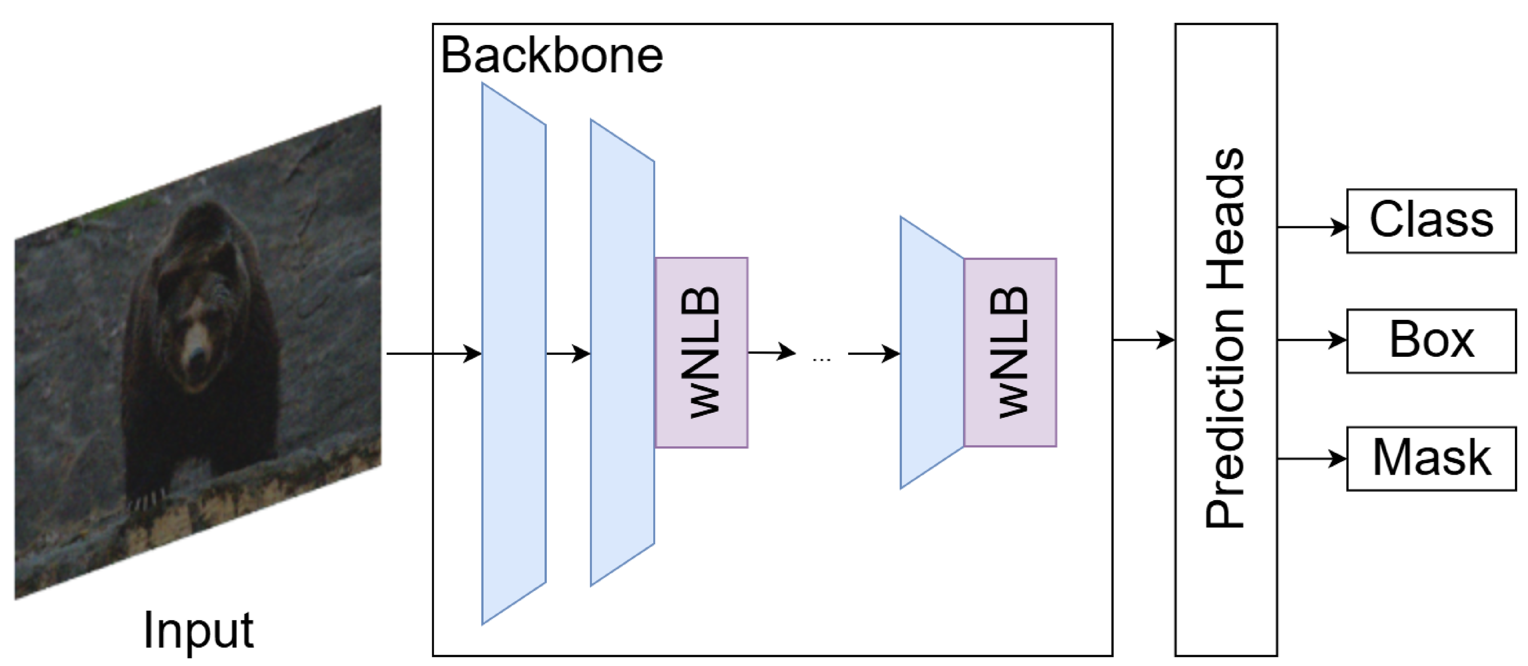

Figure 1. Generic architecture showing our proposed weighted non-local blocks added into the backbone to remove noise in the feature space. Blue blocks indicate convolutional layers.

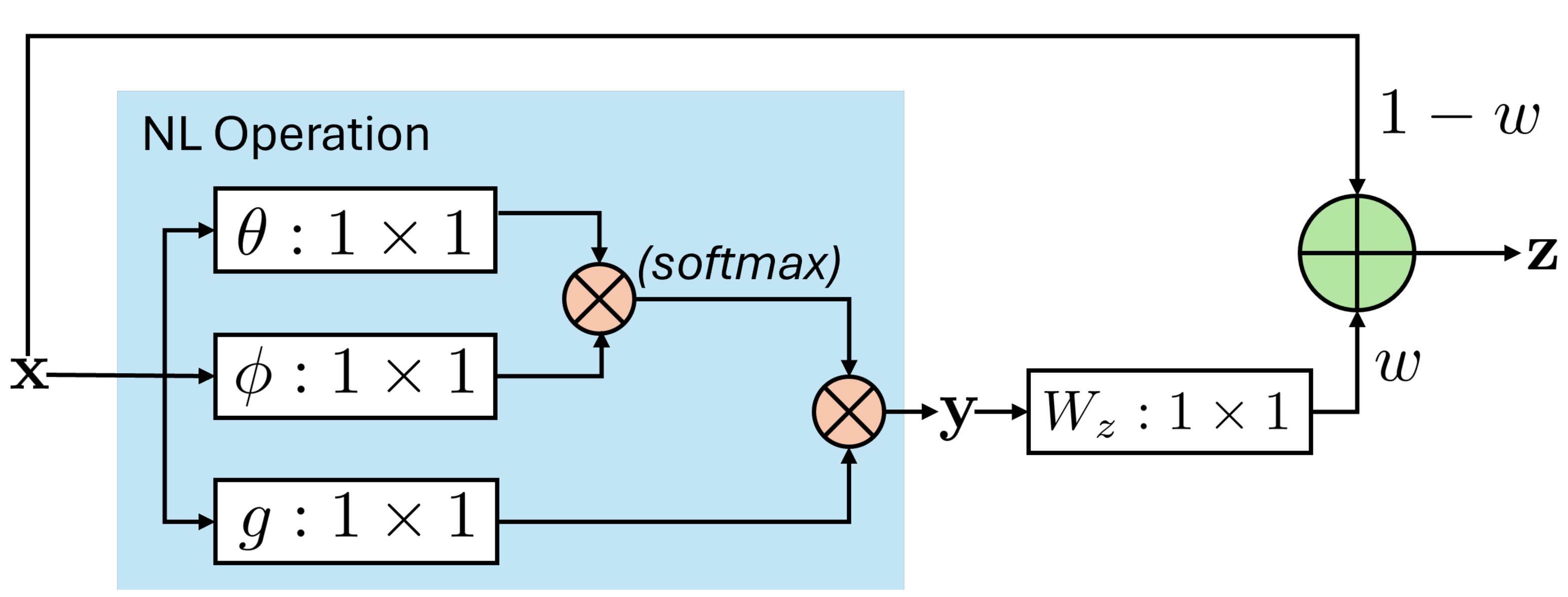

Figure 2. Our proposed weighted non-local block (wNLB) for feature denoising with learnable weight w.

Quantitative Results

Table 1. Comparison of instance segmentation methods on the synthetic low-light COCO minival dataset

Method		AP	AP$_{50}$	AP$_{75}$	AP$_S$	AP$_M$	AP$_L$
Mask R-CNN	Pre-trained	6.9	12.4	6.9	2.3	7.4	12.4
Mask R-CNN	Finetuned	15.9	28.6	15.6	4.8	15.8	27.8
Mask R-CNN	NLB	16.6	30.3	16.4	5.4	16.7	28.1
Mask R-CNN	wNLB	**16.9**	**30.7**	**16.6**	**5.6**	**17.2**	**28.9**
YOLOv8	Pre-trained	6.3	11.1	6.3	1.8	6.8	10.7
YOLOv8	Finetuned	14.3	25.6	14.3	4.0	14.0	24.1
YOLOv8	NLB	**22.1**	**37.6**	**22.2**	**7.4**	23.1	**36.6**
YOLOv8	wNLB	22.0	37.5	22.1	**7.4**	**23.2**	36.5
SOLOv2	Pre-trained	8.2	14.1	8.3	2.7	8.6	14.1
SOLOv2	Finetuned	15.0	26.5	14.9	3.9	15.0	26.5
SOLOv2	NLB	**15.8**	**27.9**	15.6	**4.1**	**15.8**	27.6
SOLOv2	wNLB	**15.8**	**27.9**	**15.8**	**4.1**	15.7	**27.8**

Table 2. Comparison of two-stage methods on the synthetic low-light COCO minival dataset

Method	AP	AP$_{50}$	AP$_{75}$	AP$_S$	AP$_M$	AP$_L$
EnlightenGAN	5.5	10.0	5.5	1.7	6.1	10.0
ZeroDCE++	5.6	10.0	5.6	1.8	6.1	9.6
AGLLNet	6.1	11.0	6.1	1.8	7.2	10.5
RetinexFormer	5.7	10.3	5.7	1.8	6.5	9.9
Ours	**16.9**	**30.7**	**16.6**	**5.6**	**17.2**	**28.9**

Qualitative assessment on a real low-light dataset

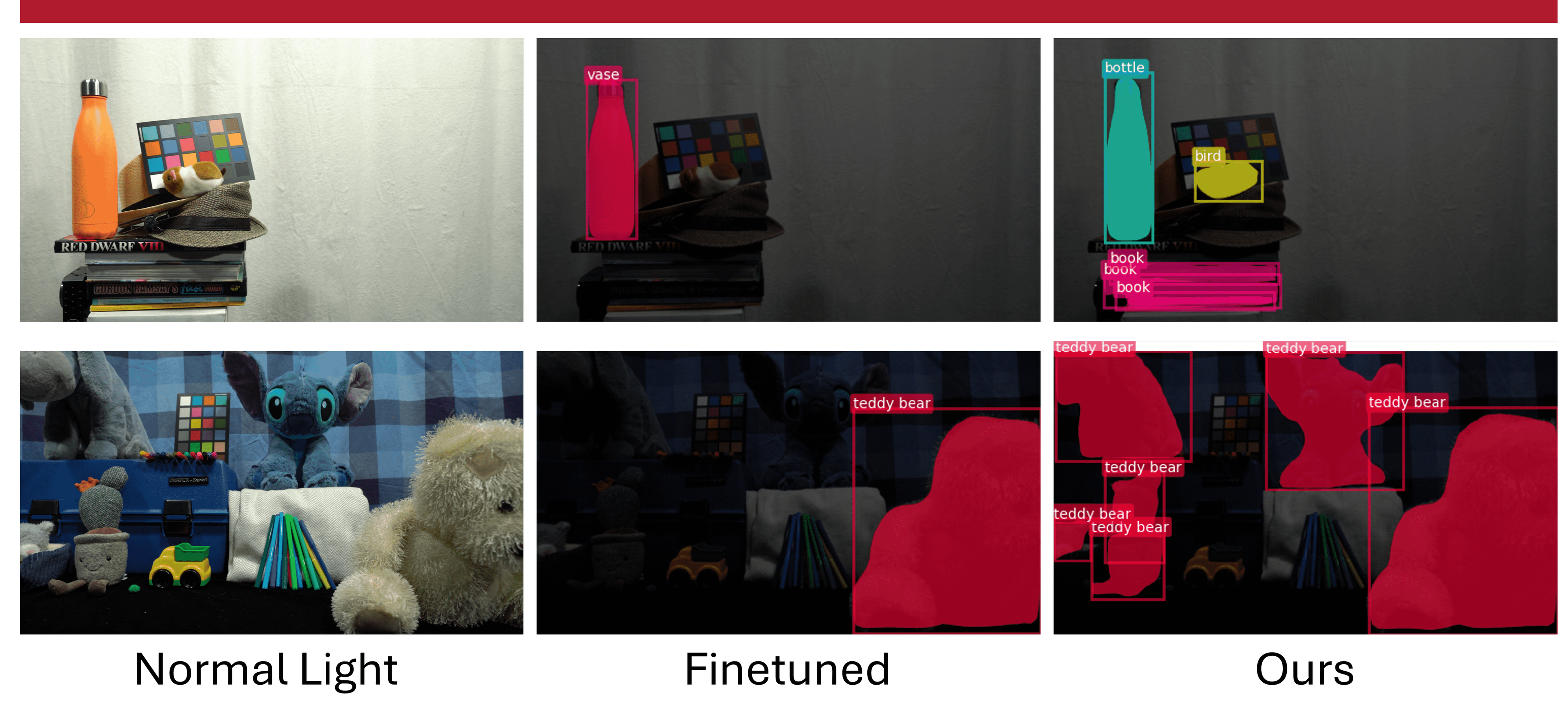

Figure 3. Visual comparison of our proposed method against the finetuned method using the Mask R-CNN [2] architecture on real low-light data from the BVI-RLV Video dataset [3].

Qualitative Results

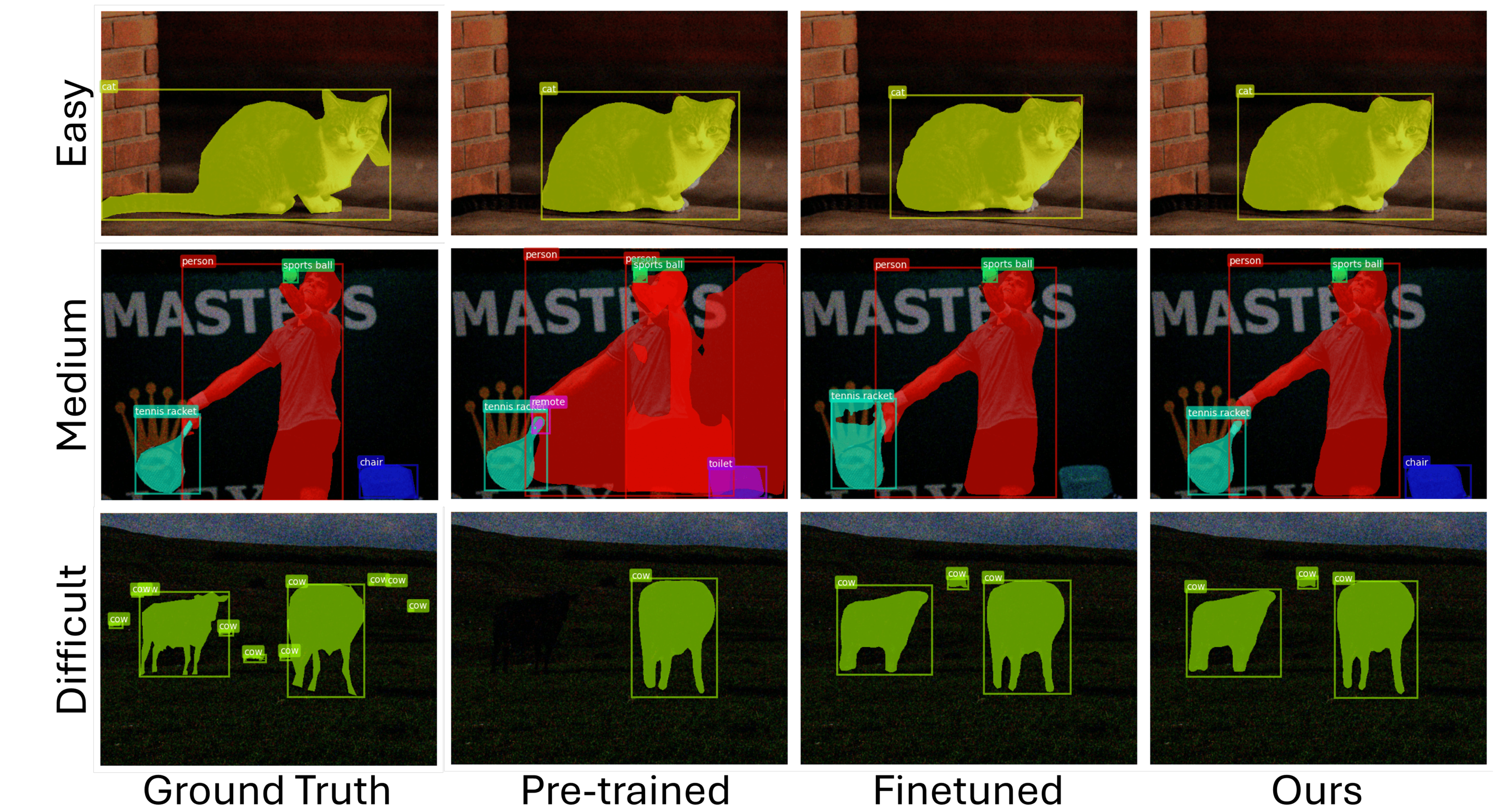

Figure 4. Visual comparison of our proposed method against pre-trained and finetuned Mask R-CNN [2] models, along with the ground truth, for cases of varying levels of difficulty.

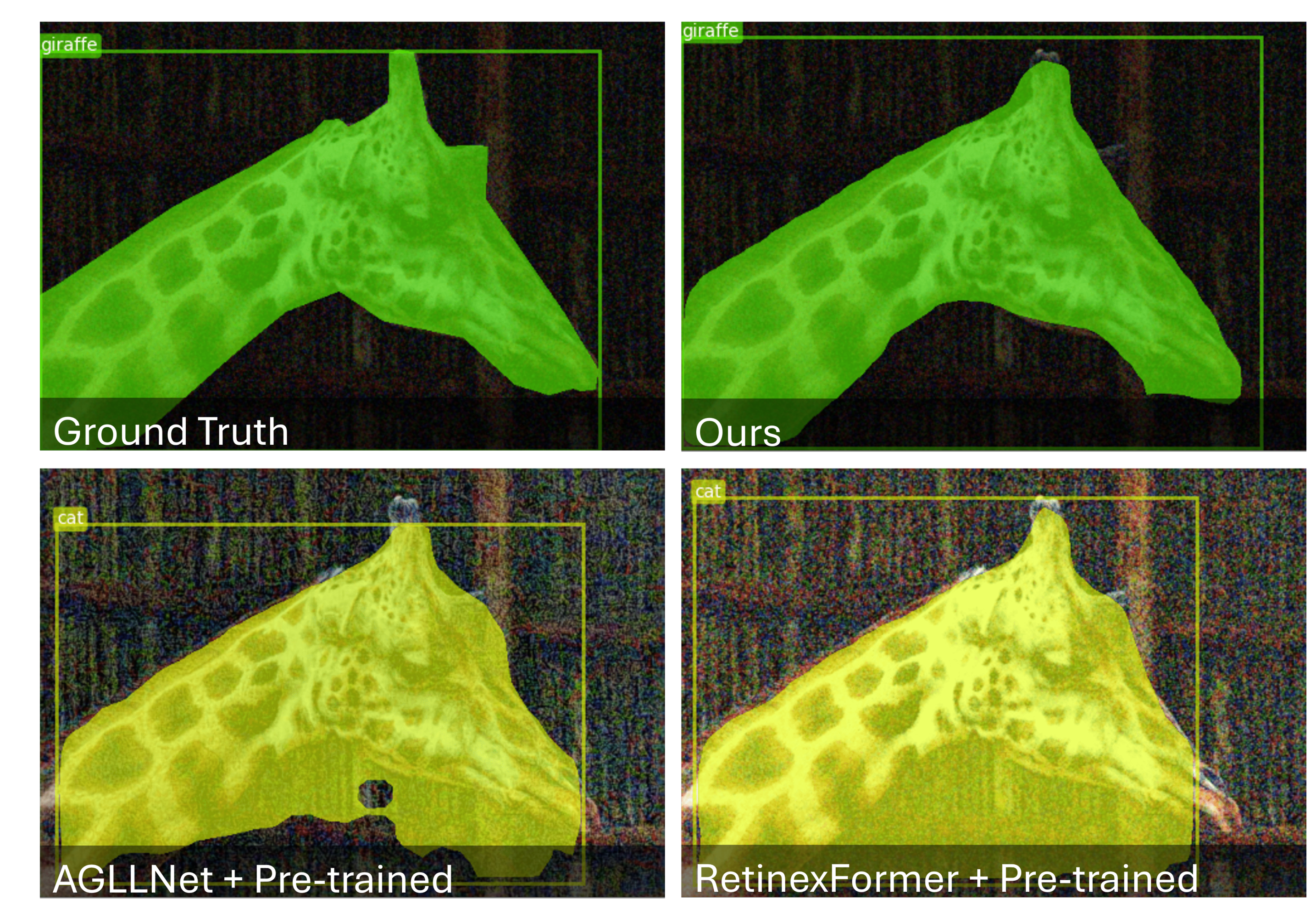

Figure 5. Visual comparison of our proposed method against two-stage methods (enhanced first then passed through a pre-trained model) using Mask R-CNN [2] as the detector.

References

[1] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[2] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2980–2988.

[3] R. Lin, N. Anantrasirichai, G. Huang, J. Lin, Q. Sun, A. Malyugina, and D. Bull, "BVI-RLV: A fully registered dataset and benchmarks for low-light video enhancement," *arXiv preprint arXiv:2407.03535*, 2024.

joanne.lin@bristol.ac.uk