UEMNet: A Lightweight General-Purpose Convolutional Neural Network for Environmental Applications

Thomas Wang Wei Hong[a], **Ningchang Gan**[b]

[a] *Environmental Research Institute, National University of Singapore, Singapore 117411* thomasww@nus.edu.sg

[b] *Department of Statistics and Data Science, National University of Singapore, Singapore 117546* e1351391@u.nus.edu

* Presenting author

1. Introduction

The rapid advancement of deep learning has significantly improved image processing capabilities, facilitating research in environmental science. For instance, convolutional neural networks (CNNs) have been employed to automate the identification of microplastics in images, enhancing efficiency. However, this progress introduces a new challenge—the demand for compact and lightweight models suitable for real-world environmental applications. Researchers in the environmental domain often lack the computational resources and expertise to deploy large-scale deep learning models, necessitating alternative lightweight solutions for image processing tasks. To address this limitation, we proposed UEMNet, a lightweight general-purpose Convolutional Neural Network designed for efficient image analysis in environmental field. The main novelties behind the UEMNet includes (1) the use of octave convolutions over standard convolutions, (2) a simple parameter-free attention module, and (3) an ingenious integration of both U-Net++ and UNet3+. We intend to evaluate UEMNET on two environmental tasks: microplastic image segmentation and remote sensing image segmentation. Although our proposed model has yet to undergo experimental validation, its theoretical design already presents several advantages over existing architectures, thus making it a promising alternative for environmental applications.

2. Related work

2.1 U-Net++ and UNet3+

U-Net++ is an upgrade of the ubiquitous U-Net [1] where the simple skip connections are replaced by nested, dense skip connections together with the use of deep supervision, which gives the U-Net++ the ability to undergo model pruning to yield a smaller and yet functional model when necessary [2].

UNet3+ is a further improvement over U-Net++ in which (1) full-scale skip connections replaced the nested, dense skip connections, (2) use of deep supervision on feature maps of different resolution from decoder network and (3) use of a new hybrid loss function for model training [3].

2.2 MobileNetV2

MobileNetV2 serves to improve upon MobileNetV1 [4] with the introduction of linear bottlenecks and inverted residuals, which is built upon the idea that feature maps inside of a CNN can actually be represented by manifolds that exist in lower-dimensional subspaces, which is well represented by the use of bottleneck layers [5].

2.3 Octave convolutions

In Octave convolutions, the input feature maps are first decomposed into a high-frequency factor and a low-frequency factor by a user pre-determined ratio, after which the low-frequency factor is then downsampled to reduce the feature map size to save on computational resources while maintaining all the low-frequency information. Information update between input and output factors of the same frequency, and information exchange between input and output factors of different frequencies occur simultaneously via the use of standard convolutions [6].

2.4 SimAM module

The Simple, Parameter-Free Attention Module (SimAM) is modelled after how the attention mechanism works in the human brain, and serves as an upgrade over existing attention modules which focuses either on the channel domain or spatial domain, corresponding to the feature-based attention and spatial-based attention mechanism in the human brain respectively [7], by integrating these two mechanisms together via the use of 3-D weights to generate weights for each neuron [8]. In fact, SimAM does not introduce additional model parameters owing to its simple design, which makes it practical and easily implementable for many scenarios.

3. Network architecture

In this section, we discussed the main rationales behind the design of our proposed model, the U-Environ-MobileV2-Net (UEMNet) for the sake of brevity. Figure 1 provides an overview of the architecture of UEMNet.

We modified the nested dense skip connections in U-Net++ to keep only the ones at the same resolution as the input image, and also to upsample and concatenate the feature maps from the deeper layers di-

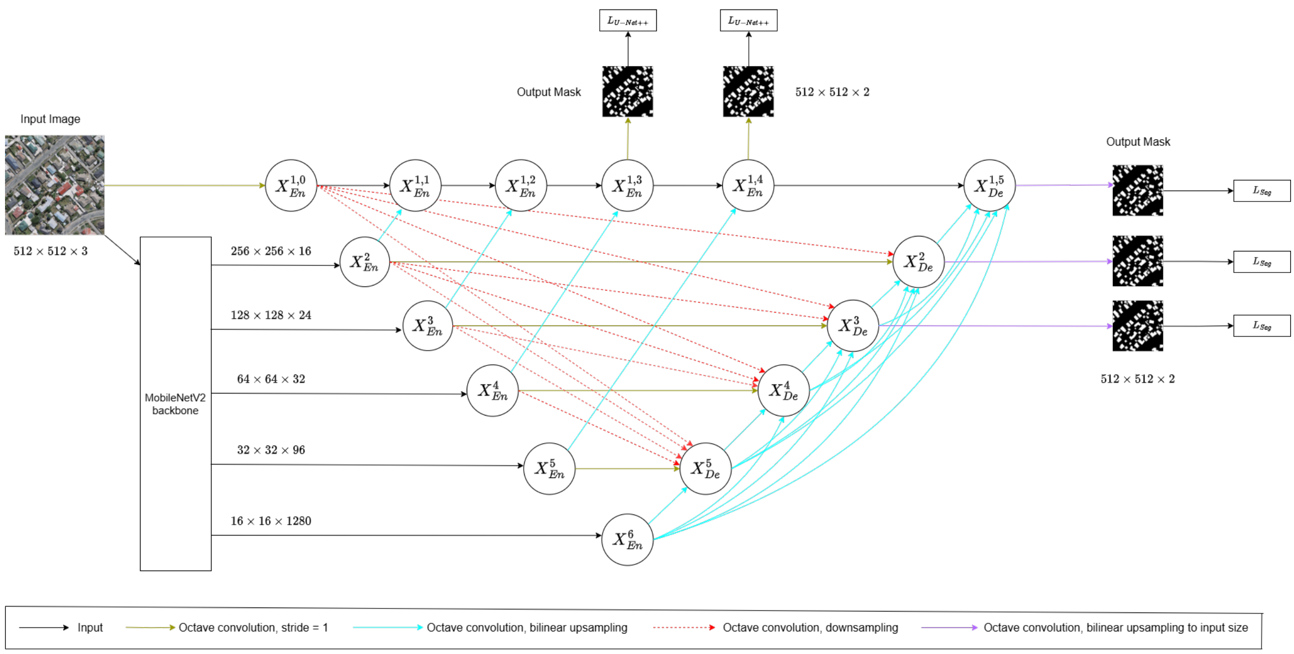

Fig. 1: An overview of the proposed UEMNet architecture

rectly to the full resolution layers so that the encoder layers can be updated directly through $L_{U-Net++}$, which is the same loss function used in [2].

In addition, we only utilize a portion of the feature maps from the higher resolution for computing the loss function instead of all resolutions as in [3], and also a portion of the feature maps from the full resolution feature maps instead of all feature maps as in [2]. This is to ensure that the updates to the weights from the loss functions will not be too rapid and cause the algorithm to diverge from the optimum loss value.

Furthermore, the SimAM is placed before every octave convolution (except for the case where the input is the input image itself) to refine the outputs from the feature maps so that the octave convolution can extract more informative features from the updated feature maps [8].

4. Future work

As a next step towards validating our findings, we intend to utilize two datasets, one from the domain of microplastics segmentation and another from the remote sensing domain.

MP-Set is a dataset of Nile Red-stained microplastics together with annotated masks which is created and introduced for the study of using U-Net with ResNet-101 encoder for the semantic segmentation of microplastics from microscopy fluorescence images [9]. It consists of a spiked and a real component, thus making it suitable for evaluating our model in both controlled and natural environments.

Global Building Dataset (GBD) is a dataset consisting of approximately 800,000 high spatial resolution (0.25 m) of diverse building styles worldwide which covers all continents of the Earth except Antarctica, hence rendering it a benchmark dataset to test the generalization ability of models in segmenting buildings from remote sensing images under all possible environments [10].

We would also be evaluating the UEMNet against some general and domain-specific models, using the Intersection over Union (IoU), precision, recall and F1 score as our evaluation metric.

The general models consist of base U-Net, U-Net with ResNet-101 encoders and DeepLabV3+ [11], while domain-specific models include U-MobileNet [12] (simplified U-Net with MobileNetV2 structure as encoders) in the domain of robotics navigation and BuildTransformer [13] (Vision Transformers (ViTs) with a dual-path structure) for building segmentation from remote sensing images.

We will publish our results in a separate journal paper, once the experiments have been completed, to demonstrate the efficacy and robustness of our proposed UEMNet.

Acknowledgments

We would like to thank our fellow colleagues, Professor Yu Liya and her team for their support, without which this work would not have been possible.

References

[1] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.

[2] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang.

Unet++: A nested u-net architecture for medical image segmentation. In Danail Stoyanov, Zeike Taylor, Gustavo Carneiro, Tanveer Syeda-Mahmood, Anne Martel, Lena Maier-Hein, João Manuel R.S. Tavares, Andrew Bradley, João Paulo Papa, Vasileios Belagiannis, Jacinto C. Nascimento, Zhi Lu, Sailesh Conjeti, Mehdi Moradi, Hayit Greenspan, and Anant Madabhushi, editors, *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 3–11, Cham, 2018. Springer International Publishing.

[3] Huimin Huang, Lanfen Lin, Ruofeng Tong, Hongjie Hu, Qiaowei Zhang, Yutaro Iwamoto, Xianhua Han, Yen-Wei Chen, and Jian Wu. Unet 3+: A full-scale connected unet for medical image segmentation. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1055–1059, May 2020.

[4] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications, 2017.

[5] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[6] Yunpeng Chen, Haoqi Fan, Bing Xu, Zhicheng Yan, Yannis Kalantidis, Marcus Rohrbach, Shuicheng Yan, and Jiashi Feng. Drop an octave: Reducing spatial redundancy in convolutional neural networks with octave convolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.

[7] Marisa Carrasco. Visual attention: The past 25 years. *Vision Research*, 51(13):1484–1525, 2011. Vision Research 50th Anniversary Issue: Part 2.

[8] Lingxiao Yang, Ru-Yuan Zhang, Lida Li, and Xiaohua Xie. Simam: A simple, parameter-free attention module for convolutional neural networks. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 11863–11874. PMLR, 18–24 Jul 2021.

[9] Ji Yeon Baek, Maria Krishna de Guzman, Ho-min Park, Sanghyeon Park, Boyeon Shin, Tanja Cirkovic Velickovic, Arnout Van Messem, and Wesley De Neve. Developing a segmentation model for microscopic images of microplastics isolated from clams. In Alberto Del Bimbo, Rita Cucchiara, Stan Sclaroff, Giovanni Maria Farinella, Tao Mei, Marco Bertini, Hugo Jair Escalante, and Roberto Vezzani, editors, *Pattern Recognition. ICPR International Workshops and Challenges*, pages 86–97, Cham, 2021. Springer International Publishing.

[10] Xin Huang, Zhen Zhang, and Jiayi Li. China's first sub-meter building footprints derived by deep learning. *Remote Sensing of Environment*, 311:114274, 2024.

[11] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation, 2018.

[12] Tao Hong, Botao Zhang, Qi'an Zhang, Sergey A. Chepinskiy, and Anton A. Zhilenkov. A terrain segmentation method based on u-mobilenet for navigation of outdoor robots. In *2022 34th Chinese Control and Decision Conference (CCDC)*, pages 5557–5562, Aug 2022.

[13] Libo Wang, Shenghui Fang, Xiaoliang Meng, and Rui Li. Building extraction with vision transformer. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–11, 2022.