AI-eXplained

Chia-Wei Chao, Daniel Winden Hwang, Hung-Wen Tsai, Shih-Hsuan Lin, Wei-Li Chen, Chun-Rong Huang, and Pau-Choo Chung
National Cheng Kung University, TAIWAN

Multi-Magnification Attention Convolutional Neural Networks

Abstract

To apply convolutional neural networks (CNNs) on high-resolution images, a common approach is to split the input image into smaller patches. However, the field-of-view is restricted by the input size. To overcome the problem, a multi-magnification attention convolutional neural network (MMA-CNN) is proposed to analyze images based on both local and global features. Our approach focuses on identifying the importance of individual features at each magnification level and is applied to pathology whole slide images (WSIs) segmentation to show its effectiveness. Several interactive figures are also developed to enhance the reader's understanding of our research.

I. Introduction

Convolutional neural networks (CNNs) are nowadays used for many image analysis tasks. Even though CNNs perform well in many cases, they struggle to deal with applications involving large-scale and high-resolution images, which typically contain hundreds of thousands of pixels. The problem of large-scale input images can be addressed by splitting the input image into multiple small patches, which are then input into the model. The model subsequently performs its assigned task based on the input image patches. However, in many applications, it is necessary to scrutinize not only the local features of the input image patches but also the global features of the input image at lower magnifications. In other words, rather than the single-magnification view used in most CNN models, it is necessary to adopt a multi-magnification approach, in which the model accesses both the global and local features of the image for the final decision.

Several methods have been proposed for imitating such multi-magnification observation processes [1], [2]. Although [2] enables the model to adaptively determine the general importance of all of the features extracted at each magnification, it is hard to ascertain the weights of the features at each magnification. Consequently, attention mechanisms [3], which allow weights of extracted features at each magnification to be adapted individually, have recently attracted a great deal of interest.

From the discussions above, it is clear that multi-scale observations and attention mechanisms provide an ideal opportunity for furthering the applications of CNNs in many fields. In this article, a multi-magnification attention convolutional neural network (MMA-CNN) is proposed to solve the aforementioned problems and applied to achieve patch-wise classification and approximated segmentation of high-resolution pathology whole slide images (WSIs). MMA-CNN combines multi-scale observations and attention mechanisms to improve the performance of CNNs, which only considers a single-magnification view of the target object.

II. MMA-CNN

The architecture of MMA-CNN is shown in Figure 1. In MMA-CNN, three modules, including the feature extraction module (FEM), feature integration and magnification attention module (FIMAM), and classification module (CM), are proposed to solve the patch-wise classification problem. The FEM contains two parallel CNNs to extract feature maps from image patches of WSIs of the high and low magnifications, respectively.

To extract larger field-of-view information of image patches of the low magnification, an ASPP block [4] is appended after the CNN to learn atrous feature maps from these image patches. The ASPP block is composed of a 1×1 convolutional layer and an atrous convolution layer. Different from general convolutions, atrous convolutions expand the field-of-views without increasing the kernel sizes and the number of parameters. In this light, two different feature maps of image patches of the WSIs with respect to two different magnifications are obtained.

To discover the important information from these two feature maps, the FIMAM is proposed. It concatenates two feature maps at first to obtain the concatenated feature maps. Then, the magnification-attention block is applied to extract attention feature maps, which serve as the inputs of the CM to classify the labels of the image patches. The magnification-attention block is constructed based on the squeezing and excitation block [3]. The concatenated feature maps are squeezed to obtain global spatial information by using global average pooling. Then, the excitation process is applied to obtain the important information of each channel. Finally, a channel-wise multiplication is applied to obtain the attention feature maps. With attention feature maps, the CM containing three convolutional

The full article with interactive content is available at https://doi.org/10.1109/MCI.2023.3277771.

Digital Object Identifier 10.1109/MCI.2023.3277771
Date of current version: 13 July 2023

Corresponding author: Pau-Choo Chung (e-mail: pcchung@ee.ncku.edu.tw).

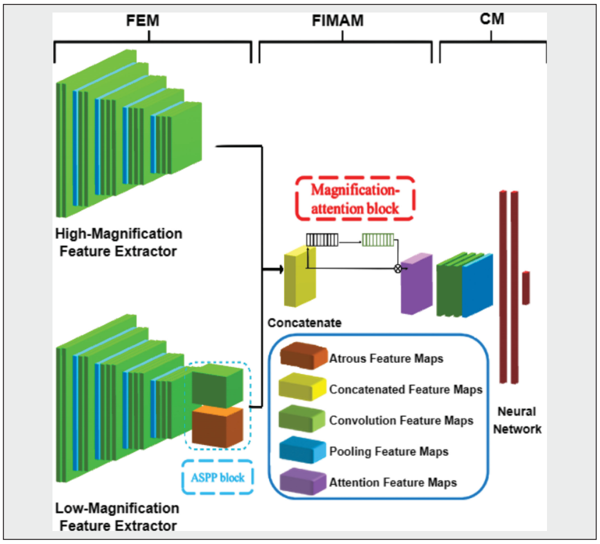

FIGURE 1 The architecture of MMA-CNN.

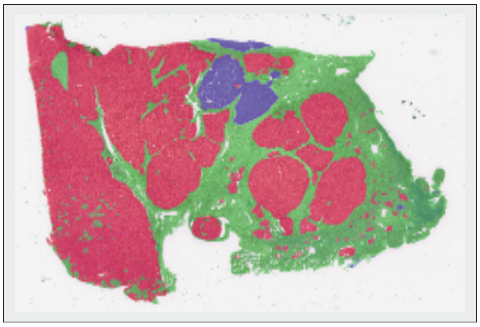

FIGURE 2 Approximated segmentation results of the tumor (red), necrosis (blue), and normal (green) regions in a stained liver WSI.

TABLE I The quantitative results of different methods.

NETWORK STRUCTURE	SCALE	TUMOR		NECROSIS	
		SENSITIVITY	IoU	SENSITIVITY	IoU
Single Network	40×	0.890	0.846	0.935	0.926
Single Network	10×	0.923	0.887	0.964	0.951
Single Network	2.5×	0.938	0.894	0.959	0.947
Parallel Network	40×&10×	0.968	0.936	0.967	0.950
Parallel Network	10×&2.5×	0.985	0.913	0.973	0.954
MMA-CNN	40×&10×	0.991	0.977	0.967	0.958
MMA-CNN	10×&2.5×	0.976	0.937	0.977	0.955

layers, a max pooling layer, and three fully connected layers with a softmax function is applied to classify the labels of the image patches of different magnifications.

III. Experimental Results

For the demonstration purposes, MMA-CNN is applied to solve the patch-wise pathology classification problem and remap the classified image patches to the pathology WSI to obtain the approximated segmentation results shown in Figure 2. The stained liver WSIs containing tumor and necrosis tissues in this study were obtained from the Leica scanner at 40 ×, 10× and 2.5× optical magnifications in National Cheng Kung University Hospital. In our experiments, 15 training WSIs, five validation WSIs, and seven testing WSIs were applied. Our method is compared with the methods that only used a single network to classify image patches of WSIs of the single magnification and the parallel networks that naively concatenate features of WSIs of different magnifications. Table I shows the quantitative results. Our method achieves the best sensitivity and IoU results because the ASPP and magnification-attention blocks help extract representative features.

IV. Conclusion

In this article, MMA-CNN is introduced, which leverages multi-magnification and attention mechanisms, to improve classification performance. By applying both global and local feature information, MMA-CNN achieves better segmentation and classification results in the stained liver WSIs compared with the competing methods using global or local features alone. In the future, MMA-CNN will be explored for potential applications besides pathology image analysis.

Acknowledgment

This work was supported in part by the National Science and Technology Council, Taiwan under Grants NSTC 111-2634-F-006-012 and NSTC 111-2628-E-006-011-MY3. We would like to thank National Center for High-performance Computing (NCHC) for providing computational and storage resources.

References

[1] W.-C. Huang et al., "Automatic HCC detection using convolutional network with multi-magnification input images," in *Proc. IEEE Int. Conf. Artif. Intell. Circuits Syst.*, 2019, pp. 194–198.

[2] H. Tokunaga, Y. Teramoto, A. Yoshizawa, and R. Bise, "Adaptive weighting multi-field-of-view CNN for semantic segmentation in pathology," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 12589–12598.

[3] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.

[4] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.