

CBAM: Convolutional Block Attention Module

Sanghyun Woo^{*1}, Jongchan Park^{*†2}, Joon-Young Lee³, and In So Kweon¹

¹ Korea Advanced Institute of Science and Technology, Daejeon, Korea
{shwoo93, iskweon77}@kai.st.ac.kr

² Lunit Inc., Seoul, Korea
jcpark@lunit.io

³ Adobe Research, San Jose, CA, USA
jolee@adobe.com

Abstract. We propose Convolutional Block Attention Module (CBAM), a simple yet effective attention module for feed-forward convolutional neural networks. Given an intermediate feature map, our module sequentially infers attention maps along two separate dimensions, channel and spatial, then the attention maps are multiplied to the input feature map for adaptive feature refinement. Because CBAM is a lightweight and general module, it can be integrated into any CNN architectures seamlessly with negligible overheads and is end-to-end trainable along with base CNNs. We validate our CBAM through extensive experiments on ImageNet-1K, MS COCO detection, and VOC 2007 detection datasets. Our experiments show consistent improvements in classification and detection performances with various models, demonstrating the wide applicability of CBAM. The code and models will be publicly available.

Keywords: Object recognition, attention mechanism, gated convolution

1 Introduction

Convolutional neural networks (CNNs) have significantly pushed the performance of vision tasks [1,2,3] based on their rich representation power. To enhance performance of CNNs, recent researches have mainly investigated three important factors of networks: depth, width, and cardinality.

From the LeNet architecture [4] to Residual-style Networks [5,6,7,8] so far, the network has become deeper for rich representation. VGGNet [9] shows that stacking blocks with the same shape gives fair results. Following the same spirit, ResNet [5] stacks the same topology of residual blocks along with skip connection to build an extremely deep architecture. GoogLeNet [10] shows that width is another important factor to improve the performance of a model. Zagoruyko and Komodakis [6] propose to increase the width of a network based on the ResNet architecture. They have shown that a 28-layer ResNet with increased

^{*}Both authors have equally contributed.

[†]The work was done while the author was at KAIST.

