Deep Learning Course - Final Project Proposal

Exploration of Attention Mechanisms for Image Classification

Netanel Madmoni & Gil Ben Or

June 23, 2023

1 Introduction

The concept of attention, introduced in [1] and later popularized by the Transformer architecture introduced in [2], is a core mechanism in state-of-the-art models for natural language processing (NLP) tasks. In recent years, more and more works in the field of computer vision are trying to incorporate this mechanism into their models, either in conjunction with a convolutional neural network (CNN) such as ResNet [3, 4], or on its own [5, 6].

The idea of integrating attention into a CNN has lead to several interesting architectural units (known as ”blocks” or ”modules”) that are designed to enhance the representational power of a CNN. Those include:

- **Convolutional Block Attention Module (CBAM)** [7], which infers attention maps along the channel and spatial dimensions. The attention maps are then applied sequentially to the input for adaptive feature refinement.
- **Bottleneck Attention Module (BAM)** [8], which is similar to CBAM, except it applies the channel attention and spatial attention modules concurrently rather than sequentially.
- **Squeeze-and-Excitation (SE) Block** [9], which ”squeezes” spatial features across channels. This is followed by an ”excitation” operation, which learns the per-channel weights. weights are applied to the feature maps to generate the output of the SE block which can be fed directly into subsequent layers of the network.
- **Efficient Channel Attention (ECA) Module** [10], Which also uses channel attention, but in a more lightweight form, using only a few parameters.

2 Proposed Work

In this project, our goals are:

1. Exploring the different modules described above.
2. Comparing the modules in terms of performance, training time and explainability.
3. Attempting to modify the internals of one or more of the modules, using different types of attention.
4. Attempting to create a hybrid attention mechanism, that combines the strengths of two or more of the modules described.

2.1 Code

We will implement this project in PyTorch code. The code we will use and modify will be taken from:

- The official PyTorch code for BAM & CBAM (by Jongchan Park on GitHub).
- The official PyTorch code for ECANet (by Banggu Wu on GitHub).
- The official PyTorch code for SE blocks (part of the PyTorch library).

2.2 Limitations

Our biggest limitation in this project is computing power. We will train and test the models on a dataset which is smaller than the one used on the papers cited above, and on significantly weaker hardware, therefore our results may vary.

References

[1] D. Bahdanau, K. Cho, and Y. Bengio, “Neural machine translation by jointly learning to align and translate,” 2016.

[2] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” 2017.

[3] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, “End-to-end object detection with transformers,” 2020.

[4] X. Wang, R. Girshick, A. Gupta, and K. He, “Non-local neural networks,” 2018.

[5] P. Ramachandran, N. Parmar, A. Vaswani, I. Bello, A. Levskaya, and J. Shlens, “Stand-alone self-attention in vision models,” 2019.

[6] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, “An image is worth 16x16 words: Transformers for image recognition at scale,” 2021.

[7] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, “Cbam: Convolutional block attention module,” 2018.

[8] J. Park, S. Woo, J.-Y. Lee, and I. S. Kweon, “Bam: Bottleneck attention module,” 2018.

[9] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, “Squeeze-and-excitation networks,” 2019.

[10] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, “Eca-net: Efficient channel attention for deep convolutional neural networks,” 2020.