Enhancing Convolutional Neural Networks with Attention Mechanisms for Medical Image Classification: A Journey from Overfitting to Data Integrity

Nina Ebensperger

December 2023

Abstract

This report delves into the development of an advanced Convolutional Neural Network (CNN) model with an integrated Convolutional Block Attention Module (CBAM) for the classification of medical images. The findings underscore the critical role of data integrity in machine learning and the effectiveness of attention mechanisms in improving model accuracy.

1 Introduction

This project embarked on the ambitious goal of employing CNNs, augmented by attention mechanisms, to classify medical images for the detection of Alzheimer's disease stages. The focus was on the incorporation of the CBAM to enhance interpretability and performance. This report chronicles the collaborative efforts and individual contributions that marked the journey from conception to the realization of this goal.

2 Individual Contributions and Algorithm Development

My individual contribution centered on constructing a CNN architecture enriched with CBAM, which introduces channel and spatial attention sequentially. Channel attention prioritizes informative features, and spatial attention concentrates on significant spatial locations within the feature maps. The interaction between these modules is governed by element-wise multiplication of the attention maps with the feature maps.

3 Attention Mechanisms in CNN

The architecture of the CNN incorporates the Convolutional Block Attention Module (CBAM) at different stages to progressively refine feature representation. The channel attention module directs the model's focus to informative features across channels, while the spatial attention module emphasizes important spatial locations in the feature maps. The equations governing the attention mechanisms are as follows:

$$F' = M_c(F) \otimes F, \tag{1}$$

$$F'' = M_s(F') \otimes F', \tag{2}$$

where F denotes the input feature maps, M_c is the channel attention map, M_s is the spatial attention map, and $\otimes$ represents element-wise multiplication. [1]

4 Channel and Spatial Attention in Convolutional Block Attention Module (CBAM)

4.1 Channel Attention

Purpose: To determine which channels (or features) in the feature map are more important.

- It starts by applying both max pooling and average pooling across the spatial dimensions of the feature map. These two pooled features capture different aspects of the feature maps.
- Both pooled features are then passed through a shared multilayer perceptron (MLP) which has one hidden layer. The MLP works as a channel descriptor, learning which channels are more important.
- The outputs from the MLP for both max and average pooled features are then combined using element-wise summation.
- A sigmoid function is applied to the combined features, resulting in a channel attention map where each value is between 0 and 1. This attention map is broadcasted and multiplied with the original feature map to scale the channels accordingly.

Purpose: To determine which channels (or features) in the feature map are more important.

- It starts by applying both max pooling and average pooling across the spatial dimensions of the feature map. These two pooled features capture different aspects of the feature maps.

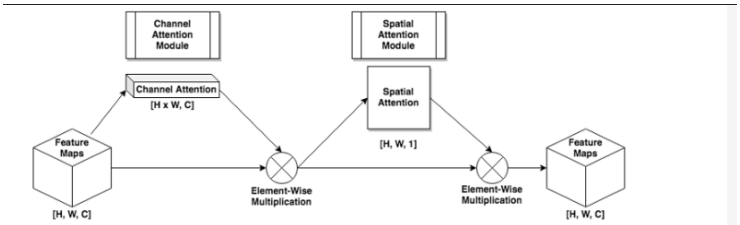

Figure 3: Block diagram of the *Convolutional Block Attention Module* (CBAM), according to Woo et al. [2018].

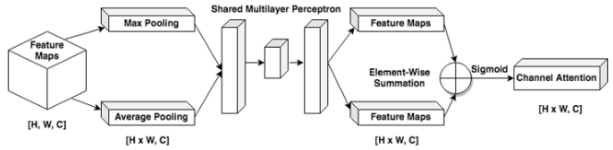

Figure 4: Block diagram of the *Channel Attention Module* of the *Convolutional Block Attention Module* (CBAM), according to Woo et al. [2018].

Figure 1: Graph illustrating the model's training progress over epochs. Adapted from [2].

- Both pooled features are then passed through a shared multilayer perceptron (MLP) which has one hidden layer. The MLP works as a channel descriptor, learning which channels are more important.
- The outputs from the MLP for both max and average pooled features are then combined using element-wise summation.
- A sigmoid function is applied to the combined features, resulting in a channel attention map where each value is between 0 and 1. This attention map is broadcasted and multiplied with the original feature map to scale the channels accordingly.

4.2 Spatial Attention

Purpose: To determine which spatial regions of the feature map should be emphasized.

- After channel attention has been applied, the model computes the average and max features across the channel dimension, which are then concatenated together.
- This concatenated feature map is then passed through a convolutional layer followed by a sigmoid function to produce a spatial attention map. This map is a single-channel 2D map where higher values indicate regions of interest.
- The spatial attention map is then used to scale the feature map after channel attention has been applied.

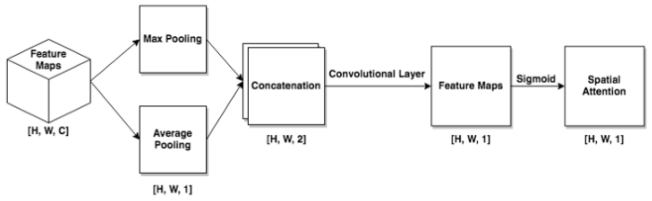

Figure 5: Block diagram of the *Spatial Attention Module* of the *Convolutional Block Attention Module* (CBAM), according to Woo et al. [2018].

Figure 2: Graph illustrating the model's training progress over epochs. Adapted from [2].

4.3 Combining Channel and Spatial Attention

- The CBAM module applies channel attention first to the input feature map, producing a feature map that has been scaled channel-wise.
- It then applies spatial attention to this channel-refined feature map, producing the final output feature map that has been refined both channel-wise and spatially.
- These attention mechanisms allow the neural network to adaptively focus on important features and suppress less relevant ones, potentially improving the performance of the network for tasks such as classification, detection, or segmentation.

5 Preprocessing and Training Process

The preprocessing incorporated data augmentation tactics to enhance the model's generalization capacity. During training, the focal loss was leveraged to address class imbalance, focusing on hard-to-classify examples. This training process was visualized through loss and accuracy graphs over epochs, which unveiled the challenge of overfitting.

6 Data Leakage Identification and Corrective Measures

An unexpected revelation of data leakage occurred late in the project, identified through anomalous model performance and data distribution insights. This prompted an immediate overhaul of the dataset structuring, ensuring patient scan slices were no longer interspersed across different data subsets.

7 Revised Results and Summary

Upon reevaluation of the model's training dynamics, signs of overfitting were identified. As shown in Figure 3, the model's training loss decreased significantly over epochs, suggesting improved learning on the training dataset. However, the validation loss plateaued and began to slightly increase, indicative of the model's diminishing ability to generalize to unseen data. Similarly, the training accuracy approached near-perfection, while the validation accuracy stagnated at a lower level, reinforcing the presence of overfitting.

The recalibrated models underwent retraining with the revised dataset, yielding new performance metrics that painted a more authentic picture of the models' capabilities. The confusion matrix, depicted in Figure 4, provided a detailed breakdown of the model's predictive accuracy across Alzheimer's disease stages, showcasing an improved balance between sensitivity and specificity.

Furthermore, Grad-CAM visualizations were employed to understand the model's focus areas during prediction. As depicted in Figure 5, the model appears to be concentrating on the skull and the space between the skull and the brain. This unexpected focus raises concerns about the interpretability of the model's learning and the potential need for further refinement to ensure that medically relevant features within the brain are being adequately considered.

This refined approach led to a more reliable assessment of the model's performance and underscored the critical role of rigorous validation in the development of machine learning models for medical applications.

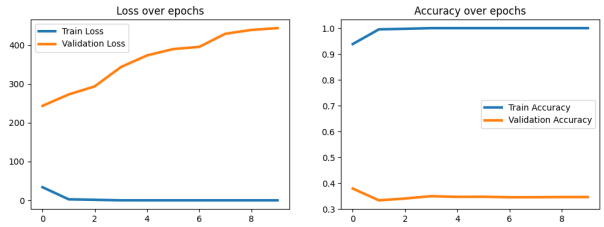

Figure 3: Loss and accuracy over epochs indicating overfitting: validation metrics do not improve in tandem with training metrics.

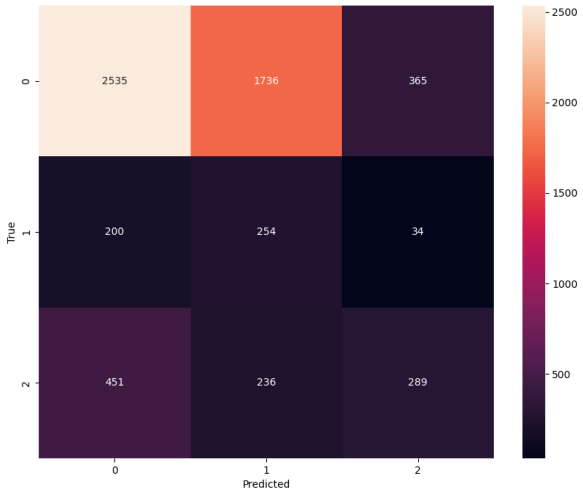

Figure 4: Confusion matrix showcasing the model's predictive accuracy across different stages of Alzheimer's disease after addressing overfitting.

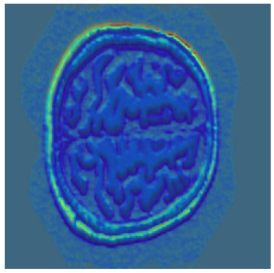

Figure 5: Grad-CAM visualization highlighting the regions of interest within a brain MRI scan that the CNN model focused on when predicting the presence of Alzheimer's disease. Areas in warmer colors indicate higher influence on the model's decision.

8 Conclusion

The project's evolution from initial model development to the resolution of data leakage encapsulates a full spectrum of machine learning development complexities. The journey reaffirmed the quintessential role of data validation in machine

learning and offered insights into the potential of attention mechanisms to elevate model performance.

9 Future Work

Despite the improvements made to the model's architecture and training process, there remain opportunities for further enhancement and exploration. For future work, we propose the following strategies:

- **Enhanced Data Augmentation:** To continue to improve the robustness of the model against overfitting, we will explore more complex data augmentation techniques. These may include geometric transformations and synthetic data generation to provide the model with a richer variety of training examples.
- **Dropout Layer Integration:** Dropout layers will be introduced at strategic points within the network architecture to prevent co-adaptation of neurons and encourage individual feature detection, reducing overfitting and improving model generalization.
- **Early Stopping Implementation:** We plan to implement early stopping criteria during training. By monitoring the validation loss and stopping the training when it begins to increase, we can prevent the model from learning noise and non-generalizable patterns in the training data.
- **Attention Mechanism Refinement:** The Grad-CAM results have prompted us to refine the attention mechanisms further. We aim to ensure that the model focuses on medically relevant features within the brain rather than extraneous regions such as the skull.
- **Hyperparameter Optimization:** We will employ rigorous hyperparameter tuning methods, potentially using automated techniques such as Bayesian optimization, to find the optimal settings for our model and improve its performance.

By addressing these areas, we anticipate enhancing the model's accuracy and reliability, pushing the boundaries of what is achievable with attention-augmented convolutional neural networks in medical image analysis.

While only accounting for the final codes submitted and ignoring packages used. My calculation yields an estimate of 85

References

[1] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. CBAM: Convolutional Block Attention Module. *arXiv preprint arXiv:1807.06521*, 2018.

[2] Tiago Gonçalves, Isabel Rio-Torto, Luís F. Teixeira, and Jaime S. Cardoso. A survey on attention mechanisms for medical applications: are we moving towards better algorithms? *arXiv preprint arXiv:2204.12406*, 2022.