Continual Learning for Object Classification: Consolidation and Reconsolidation

Daniel Turner, Pedro J.S. Cardoso, João M.F. Rodrigues

LARSyS & Instituto Superior de Engenharia, Universidade do Algarve, Faro, Portugal

Abstract

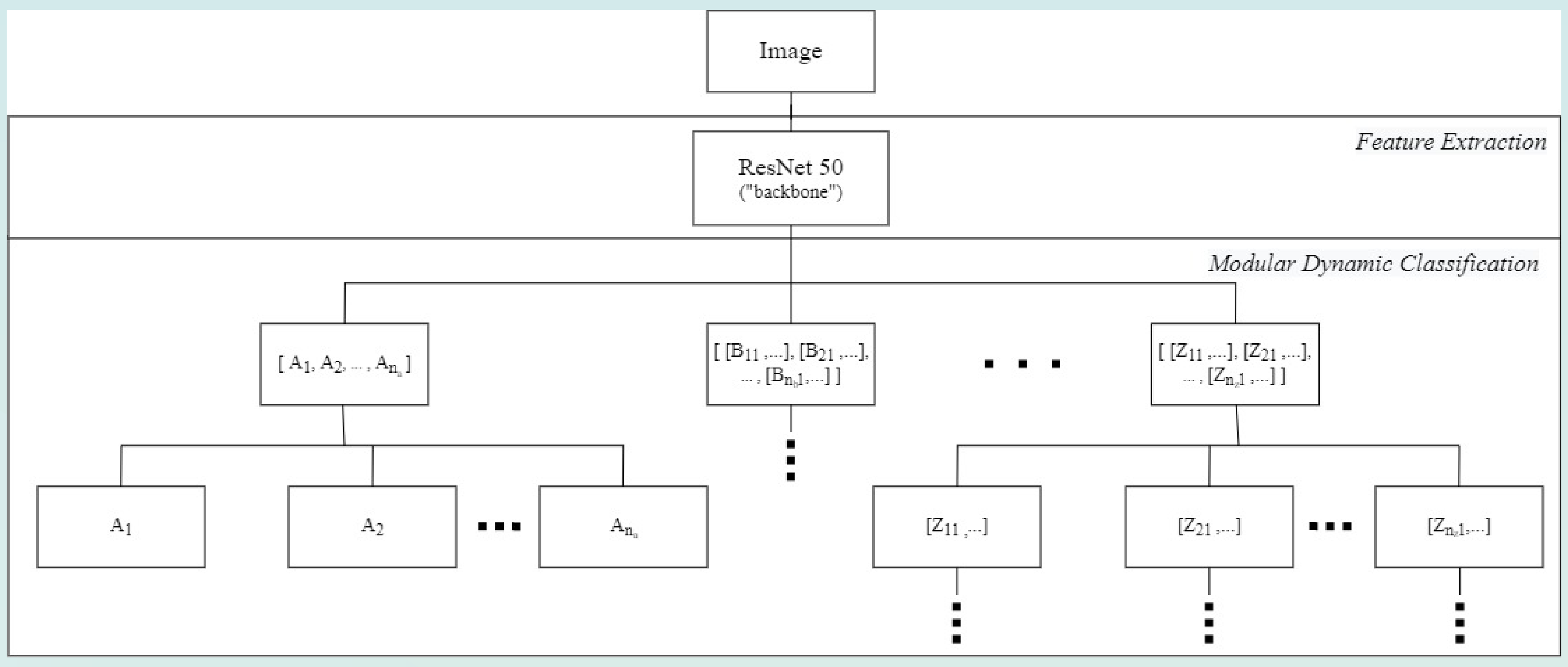
To achieve "consolidation", short-term memory's conversion to long-term memory requires the passage of time. This consolidation can occur at many organizational levels in the brain, e.g., simplifying, when two neurons repeatedly fire at the same time, they become more likely to fire together in the future. Eventually, these two neurons will become sensitized to one another. The brain progressively creates more of these connections when new information or experiences are presented. However, just because a memory has been consolidated it does not mean that it cannot be lost. Literature shows that memories often need to be reconsolidated once they have been recalled. In deep artificial neural networks, there is the tendency to forget previously learned information completely and abruptly upon learning new information, usually called catastrophic forgetting. We propose a Deep Modular Dynamic Neural Network (MDNN) based on “consolidation” and “reconsolidation” principles, capable of learning new information and mitigating the catastrophic forgetting issue. MDNN is divided in two blocks: (a) the feature extraction block, based on a ResNet50, and (b) the modular dynamic classification block, made up of modular sub-networks that progressively grow in a tree shape and re-arranges themselves as they continuously learn. Once one of these branches learn an object then it is “consolidated”. If a new object is assigned to a branch that already has consolidated objects, then all the existing objects in that branch are “reconsolidated”, this process occurs on the fly. The network presents state of the art results on CORe50 dataset. [LARSyS - FCT Project UIDB/50009/2020]

Introduction

While the challenge of teaching machines how to distinguish between sets of objects has been solved with very impressive results, the challenge of teaching them how to recognise new objects on top of those already known is still very much unsolved. Learning to recognize new objects may be a simple task for a human, but the implementation of this in machines proves to be very difficult, at least while using current methods. The go-to approaches for the first part of the problem (teaching a machine to recognize an initial set of objects) are typically based on the use of neural-networks. So intuitively, the solution to the next step (teaching a machine new objects AKA Continual Learning) would be to elaborate on the existing solutions for the first part, but the problem with this is that the trained weights in neural-networks are extremely fragile meaning that any tweaking can severely damage their capacity to perform the functions they were trained for. This phenomenon is known as Catastrophic Forgetting.

The Modular Dynamic Neural Network

For the first part of the network, (a) feature extraction, we used a pre-trained ResNet50 as a backbone, which has been shown to yield exceptional results in image classification. This feature extraction block is used only to ex-tract generic “low-level” features.

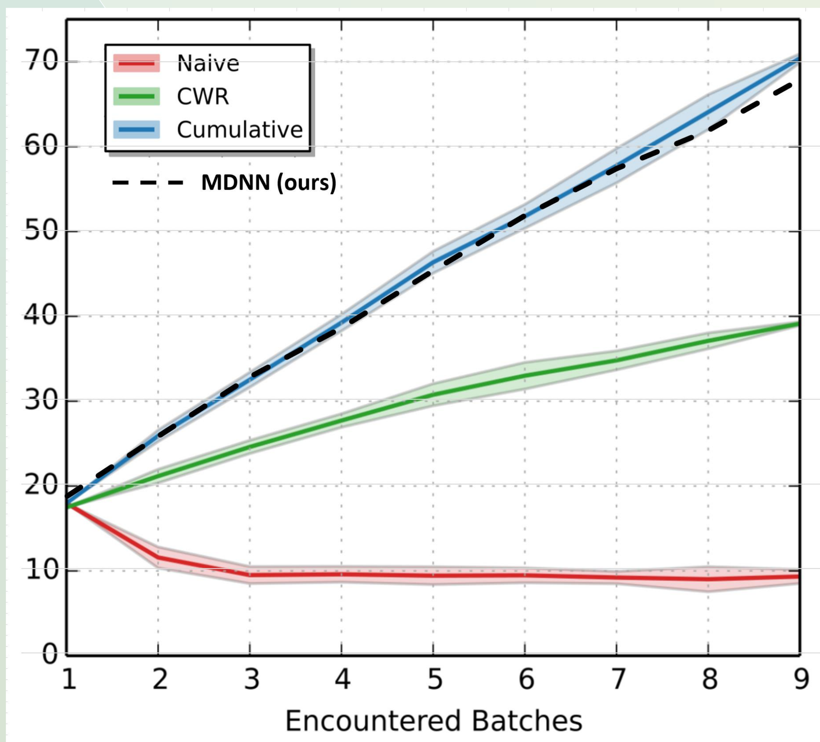


The second part of the network, the (b) modular dynamic classification block, is comprised of multiple small modules. These modules are responsible for classifying specific classes or groups of classes that, as new information is learned, are automatically divided into groups of modules and sub-modules based on their class’s similarities.

The modules which contain their own sub-modules are designated as node modules and they contain one binary classifier for each direct child module. The modules with no children are designated as endpoint modules and contain only data obtained during the training process.

Tests and Results

Test Results show very promising results when compared with those obtained on the CORe50 dataset. The displayed results correspond to the accuracy over time where the first step consists of 10 classes and then 5 more classes are added incrementally until reaching a total of 50.



CWR is the benchmark we are comparing our results with and cumulative represents a kind of upper-bound target which we are essentially reaching. The cumulative result was achieved without any CL restrictions and for our method to be so close is very positive.

Conclusions

We have presented a proof-of-concept of an image classification framework, capable of learning new classes while maintaining knowledge of previous ones. The structure of the framework is based on a modular network of smaller artificial neural networks that get added to the network when new classes are learned. This way, other parts of the network are left untouched which allows them to retain their knowledge.

Acknowledgments

This work was supported by the Portuguese Foundation for Science and Technology (FCT), project LARSyS - FCT Project UIDB/50009/2020.