Improving Performance in Neural Networks by Dendrite-Activated Connection

Carlo Metta[1], Marco Fantozzi, Andrea Papini [2], Gianluca Amato [3], Matteo Bergamaschi [4], Silvia Giulia Galfrè [5], Alessandro Marchetti[*,3], Michelangelo Vegliò [3], Maurizio Parton [3], and Francesco Morandin [6]

[1] ISTI-CNR Pisa, Italy

[2] Scuola Normale Superiore, Pisa, Italy

[3] University of Chieti-Pescara, Italy

[4] University of Padova, Italy

[5] University of Pisa, Italy

[6] University of Parma, Italy

Abstract: In artificial neural networks, computational units typically compute a linear combination of their inputs and then apply a nonlinear filter, often a ReLU, shifted by some bias. If the inputs come from other units, they have already been filtered with their own biases. In a layer, multiple units share the same inputs, and each input is filtered with a unique bias. This results in output values based on *shared* input biases rather than individually optimized ones. To address this issue, we introduce DAC, a new computational unit that incorporates preactivation and multiple biases. This design allows input signals to undergo independent nonlinear filtering before the linear combination.

In this short note, we sketch the design of this new computational unit. Full theoretical support and empirical evidence, suggesting that DAC could be an improved design for the basic computational unit in neural networks, can be found in Metta *et al.*, 2023. Code at `https://github.com/CuriosAI/dac-dev`

Keywords: preactivation, multi-bias, ResNet, dendritic neural model.

1 Introduction

Historically the structure of the perceptron, the artificial neural network's fundamental computational unit, has rarely been questioned. The biological inspiration is straightforward: input signals from the dendrites are accumulated

[*] Alessandro Marchetti is a PhD student enrolled in the National PhD in Artificial Intelligence, XXXVII cycle, course on Health and life sciences, organized by Campus Bio Medico University of Rome.

at the soma (with a linear combination), and if the result is above the activation threshold (that is, the opposite of some bias) there is a nonlinear reaction, as the neuron fires along the axon (with the activation function).

In time the early sigmoid activation function was replaced by ReLU and variants, and this has brought us to the current situation in which most units output their signal through a nonlinear activation function which effectively destroys some information. In fact, ReLU is not invertible, as it collapses to zero all negative values. Though some of its variants may be formally invertible (ELU Clevert *et al.*, 2016 for example), they overall perform in a way very similar to ReLU. This suggests that their way of compressing negative values leads to the same general properties of the latter.

In this note, we describe a radical rethinking of the standard computational unit, where the output brings its full, uncorrupted information to the next units, and only at this point is the activation function applied, with biases specialized for each unit. From the biological point of view, this is like having the activation at the dendrites instead of at the base of the axon. Thus, we call the new unit 'DAC', for 'Dendrite-Activated Connection'.

An extended version of this note, with implementation details, an efficiency analysis in terms of parameters and FLOPs, empirical evidence that DAC provides several benefits with respect to standard units, a theoretical analysis including a universal approximation theorem, and more, can be found in the full paper Metta *et al.*, 2023.

2 From standard to DAC units

To describe this paper's idea, we look at a neural network as a directed acyclic computational graph. We denote the set of its nodes by I. If $i \in I$ is a node, we denote its parents (in-neighbors) by $I_i \subset I$. In the *standard model* for computational units in a neural network, a bias b_i, a set of weights $w_{i,j}$ for $j \in I_i$ and a nonlinearity φ_i are associated with every node i. In this paper, φ_i is always $\varphi = \mathrm{ReLU}$.

Standard model network flow involves updating node i's value y_i using a nonlinear filter, *activation* $\circ$ *bias*, applied to some information linearly aggregated from node i's parents:

$$\begin{cases} z_i = \sum_{j \in I_i} w_{i,j} y_j & \text{linear aggregation} \\ y_i = \varphi(b_i + z_i) & \text{nonlinear filter} \end{cases} \tag{1}$$

Figure 1 exhibits this point of view, emphasizing that parent nodes (white boxes) are themselves filtered with biases and ReLU. For each parent node

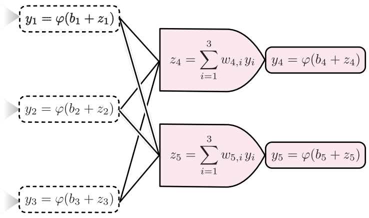

Figure 1: Standard network example. A fully connected layer with 3 input units and 2 computational units. The set of in-neighbors of nodes 4 and 5 is $I_4 = I_5 = \{1,2,3\}$. Bullets and rectangles represent linear aggregation and nonlinear filters (1), respectively. Units 4 and 5 must share the same biases b_1, b_2, b_3 in their inputs, potentially causing outputs y_4 and y_5 to be based on deteriorated information.

$i = 1,2,3$, the bias b_i is uniquely determined. Since children nodes $j = 4,5$ cannot access z_i directly, they must use the filtered versions y_i, sharing the way they are filtered. We refer to this as *shared* biases and argue that it could cause information degradation.

The direct solution to this problem is to apply the nonlinear filter with *non-shared* biases at the input of the unit, before the linear aggregation. We investigate this idea, by studying a new computational unit briefly described as *linearity* $\circ$ (*activation* $\circ$ *non-shared biases*):

$$\begin{cases} y_{i,j} = \varphi(b_{i,j} + z_j) & \text{nonlinear filter} \\ z_i = \sum_{j \in I_i} w_{i,j}\, y_{i,j} & \text{linear aggregation} \end{cases} \tag{2}$$

The artificial neuron described by (2) reflects a recent shift in the understanding of the biological neuron towards a model that incorporates *active* dendrites Larkum, 2022; Magee, 2000. Active dendrites perform a local nonlinear signal modulation before integration at the soma level. Since the biases $b_{i,j}$ in (2) can depend on both input and output nodes, that is, on the edges of the graph, and since these edges correspond to dendrites in biological neurons, we call this new computational unit a *Dendrite-Activated Connection* (DAC) unit. See Metta *et al.*, 2023 for more details on biological inspiration.

One can view DAC as a preactivated unit with multiple *non-shared* input biases, meaning that DAC units sharing the same inputs can filter them with different nonlinearity thresholds, see Figure 2.

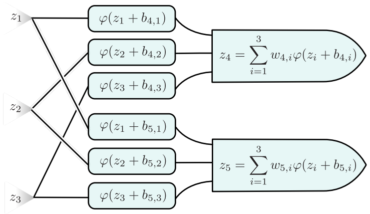

Figure 2: DAC network example: the network from Figure 1 with standard units replaced by DAC units. The input biases b_1, b_2, b_3, contributing to the output values z_4 and z_5, now depend also on the output nodes $4, 5$. We call this feature *non-shared* biases because it allows DAC units sharing the same input to use different (*non-shared*) thresholds instead of a single (*shared*) input bias.

Acknowledgements. Computational resources by CLAI laboratory of Chieti-Pescara. The authors wish to thank Rosa Gini for her important intellectual contribution.

References

CLEVERT, DJORK-ARNÃ©, UNTERTHINER, THOMAS, & HOCHREITER, SEPP. 2016 (Feb.). *Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs).* arXiv:1511.07289 [cs].

LARKUM, MATTHEW E. 2022. Are Dendrites Conceptually Useful? *Neuroscience*, **489**(May), 4–14.

MAGEE, JEFFREY C. 2000. Dendritic integration of excitatory synaptic input. *Nature Reviews Neuroscience*, **1**(3), 181–190.

METTA, CARLO, FANTOZZI, MARCO, PAPINI, ANDREA, AMATO, GIANLUCA, BERGAMASCHI, MATTEO, GALFRÈ, SILVIA GIULIA, MARCHETTI, ALESSANDRO, VEGLIÒ, MICHELANGELO, PARTON, MAURIZIO, & MORANDIN, FRANCESCO. 2023. *Improving Performance in Neural Networks by Dendrites-Activated Connections*. `https://arxiv.org/abs/2301.00924`.