Final Project Report

Group 5: Marc Artero, Diego Hernández, Bernat Medina, and Pau Monserrat

Universitat Autònoma de Barcelona, Barcelona, Spain

1 Introduction

This report serves as supplementary material to our final project presentation. Its purpose is to detail the experiments that were excluded for the sake of clarity and to address the technical concerns raised during the oral defense.

2 Omitted Experiments

We conducted several experiments that were eventually omitted from the presentation: the Fully Connected (FC), Attention, Residual Connections, and Shuffle Layers experiments. In the following subsection, we provide a detailed overview of the FC experiment. Meanwhile, in Section 3, we offer a justification for why the Attention and Residual Connections experiments failed to outperform our proposed baseline, despite these modules being theoretically orthogonal and typically expected to provide additive gains to model performance.

2.1 FC experiment

In the early stages of the assignment, we analyzed the impact of modifying the number of units in the first FC layer of our architecture. As shown in Table 1, reducing the unit count from our baseline of 64 (see Fig. 1) proved detrimental to accuracy, though it yielded better efficiency metrics.

Table 1. Results of the FC experiment (on the amount of units).

Units	Parameters ↓	Accuracy ↑	Efficiency ↑	Distance ↓
16	0.32 M	29.3 %	0.09	3.32
32	0.58 M	45.0 %	0.08	5.78
64	1.08 M	70.1 %	0.07	10.78
128	2.08 M	71.3 %	0.03	20.81
256	4.09 M	69.2 %	0.02	40.90

Given these results, we investigated the use of Global Average Pooling (GAP) as a way to eliminate all units while combining the spatial information from the feature maps. This provided a 2.6 % accuracy improvement while simultaneously removing over 93% of the total parameters (Table 2). Because GAP essentially made the analysis of the FC layer obsolete, we omitted the experiment to focus on the more effective GAP-based approach, which streamlined the project narrative.

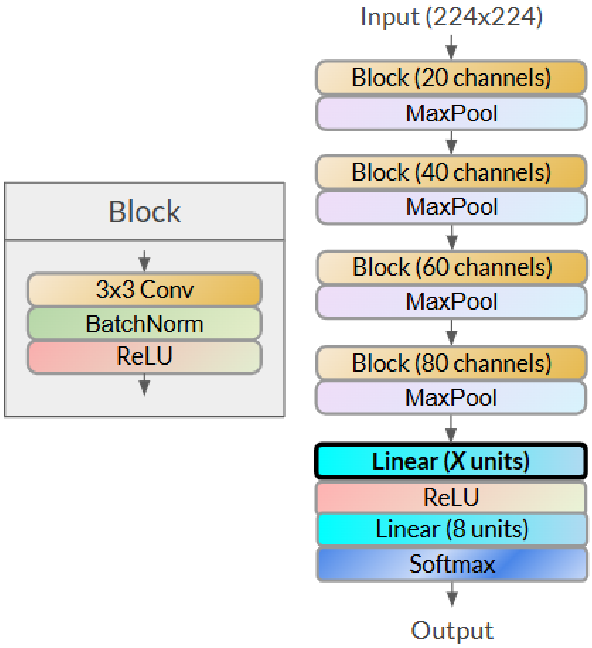

Fig. 1. Network diagram in the FC experiment.

Table 2. Results obtained when replacing the FC layer with Global Average Pooling.

Parameters ↓	Accuracy ↑	Efficiency ↑	Distance ↓
73.6 K	72.7 %	0.99	0.79

3 Analysis of Sub-Optimal Architectures

This section addresses why the introduction of Attention and Residual Connections did not yield the expected performance gains in our specific context.

3.1 Attention

Following the discussion during the defense, we hypothesized the suboptimal performance of the Attention-based model was primarily due to an unoptimized training configuration. To validate this, we integrated Attention (see Fig. 2) into our final model, optimized via Hyperparameter and Neural Architecture Search.

As presented in Table 3, under these optimized conditions, the Attention module yields a marginal accuracy improvement of 0.1 %. Nonetheless, the 36.8 % increase in parameter count leads to a degradation in both efficiency metrics. Consequently, while these results confirm that Attention is not inherently detrimental to classification performance (as pointed out during the defense), the gain is insufficient to justify the substantial increase in computational complexity.

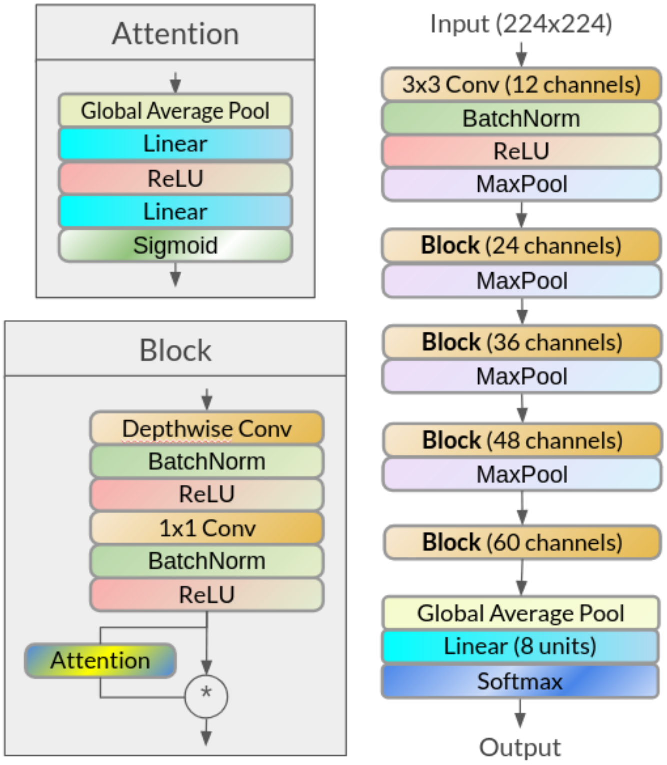

Fig. 2. Network diagram in the Attention experiment.

Table 3. Results obtained with and without the Attention module.

Attention	**Parameters ↓**	**Accuracy ↑**	**Efficiency ↑**	**Distance ↓**
×	8.26 K	81.5 %	9.86	0.20
✓	11.3 K	81.6 %	7.22	0.22

3.2 Residual Connections

The Residual Connections experiment yielded 76.1% accuracy. In our implementation, the blocks involved a change in the number of channels between the input and output. Because a direct identity addition was impossible due to this mismatch, we used a 1×1 convolution for channel expansion, as shown in Fig. 3.

We believe that these 1×1 convolutions hindered performance by distorting the identity mapping. Instead of providing a shortcut for gradient flow, the transformation introduced additional complexity that the model was unable to navigate effectively. Consequently, the skip connections likely injected noise rather than facilitating the learning of residual functions, leading to the observed decrease in accuracy compared to the baseline.

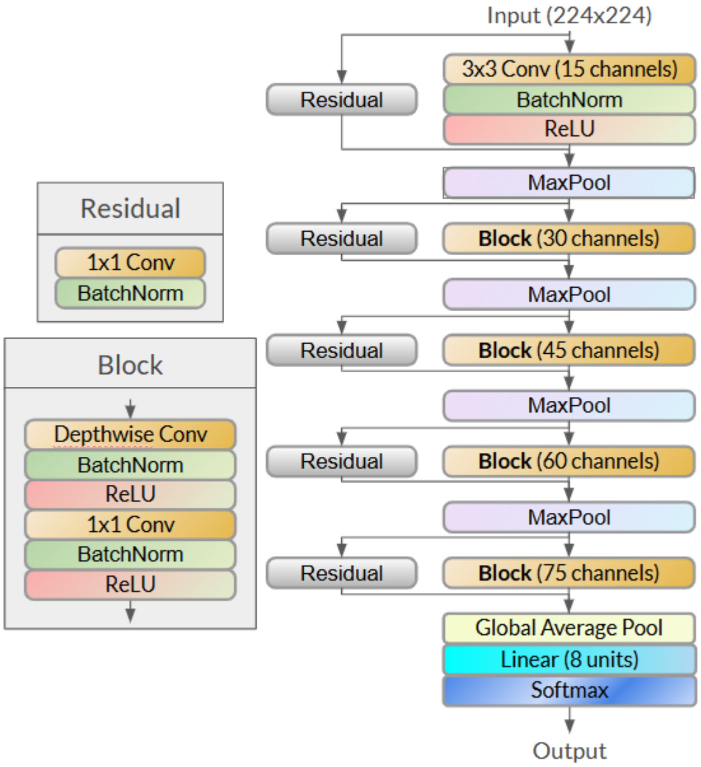

Fig. 3. Network diagram in the Residual Connections experiment.

To mitigate this, future implementations could utilize concatenative skip connections to preserve a *pure* identity mapping by stacking input features with the block's output. Unlike expansion convolutions, which transform the skip signal, concatenation allows the original information to flow forward unchanged. To manage the resulting increase in dimensionality, a 1×1 convolution could act as a bottleneck after the concatenation. Relocating the transformation from the shortcut path to the integration stage allows the model to optimize combined feature representation without corrupting the fundamental identity flow.

4 Conclusion

This project provided a comprehensive journey through the foundations of Deep Learning, from traditional handcrafted feature extraction to the design of sophisticated CNNs. Through iterative experimentation, we learned that architectural complexity does not always equate to better performance, especially when dealing with data scarcity and efficiency constraints. Mastering these basics has been essential for understanding the critical trade-offs between accuracy, parameter count, and generalization in modern Computer Vision.