Adaptive Group Normalization: Leveraging Group Normalization in   
Deep Learning Architectures using adaptive methods.  
Future Research Endeavor

Yair Smadar and Dr. Assaf Hoogi

Ariel University

1. Abstract

Deep learning, a subset of machine learning, has seen rapid advancements, substantially driven by improvements in training deep neural networks. One such vital improvement is normalization techniques, with Group Normalization (GN) [1] emerging as a robust alternative to Batch Normalization (BN) and Layer Normalization (LN). Despite GN's potential, it has some limitations, notably its static grouping of channels, potentially leading to sub-optimal performance in certain scenarios. This research proposal aims to investigate Adaptive Group Normalization (AGN), a novel approach that seeks to optimize GN by adaptively selecting groups of channels that are semantically similar according to their feature representations.

**2. Motivation**

The motivation for this research arises from the need to address the limitations of GN, particularly its lack of adaptability in grouping channels. While GN has proven beneficial in scenarios with smaller batch sizes, its fixed group division may not always be ideal for all tasks or network architectures. It is hypothesized that by allowing groups to be formed adaptively based on the closeness in meaning of their feature representations, AGN could offer a more flexible and potentially more efficient approach to normalization. Such an approach could significantly enhance the training of deep neural networks, leading to improved model performance and generalization. Furthermore, given the ubiquity of deep learning applications, advancements in normalization techniques could have wide-ranging impacts across many domains, including computer vision, natural language processing, and beyond**.**

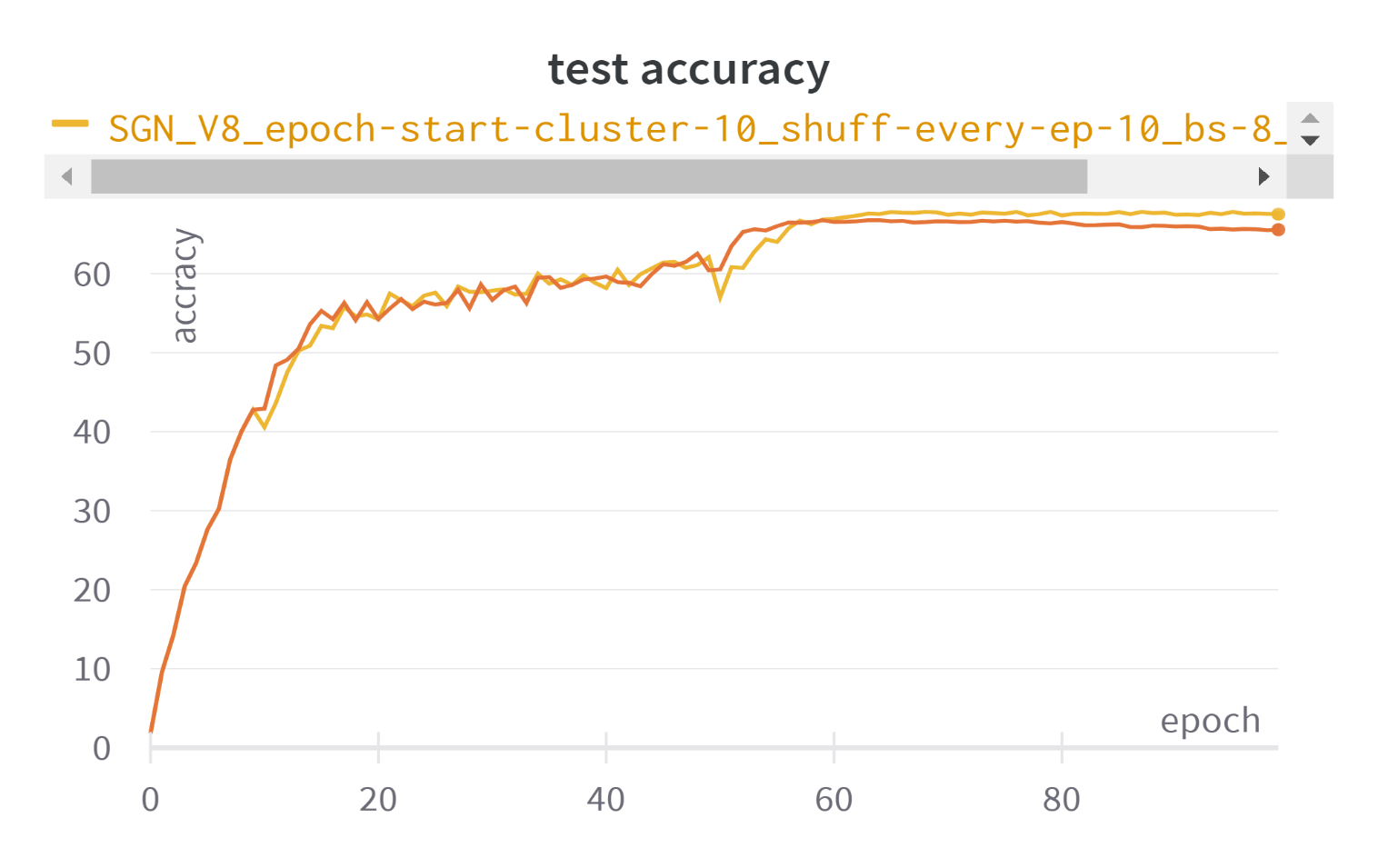
**3. Proposed Methodology:**

Our proposed research methodology involves a novel algorithm for implementing Adaptive Group Normalization (AGN). This algorithm enhances Group Normalization by adaptively selecting channels that are semantically similar based on their feature representations.   
The algorithm follows these steps:

1. **Channel Statistics Calculation:** For every channel in the batch, calculate the mean and variance. These statistics represent the primary attributes of each channel's feature distribution.
2. **Harmonic Mean Computation:** Compute the harmonic mean of the derived mean and variance for each channel. The harmonic mean, compared to other types of averages, gives a more robust measure that emphasizes the interdependence of the mean and variance.
3. **Histogram Creation and Group Division:** Arrange the computed harmonic means in a histogram. This visualization provides a clearer picture of the distribution of the channels feature representations. Following this, divide the histogram into distinct groups. The division of these groups will define the channel order for Group Normalization. This step essentially groups channels that are close in their variance and mean, under the hypothesis that these channels are likely to be semantically similar.
4. **Channel Rearrangement:** After the group normalization process, rearrange the channels back to their original order. This step ensures that the original structure of the data is preserved for subsequent layers in the network.
5. **Periodic Channel Mixing and Regular Re-mixing:** Perform the channel mixing process every few epochs, initially only for the first batch, while maintaining the mixing order for the subsequent batches. As the weights of the network converge over time, perform additional re-mixing to adapt to potential changes in the feature representations.

This methodology offers a systematic approach to implementing AGN, providing a more adaptive and potentially efficient alternative to conventional normalization techniques. It will be thoroughly evaluated through a series of experiments on various deep learning tasks and network architectures, with the results expected to shed light on the effectiveness and practical implications of AGN.

4. Results  
In Table 1, we present the empirical findings of our comparison between AGN and GN on the ResNet-50 architecture using the CIFAR-100 dataset. The table clearly delineates the superior performance of AGN across several key metrics. These findings validate our hypothesis about the efficacy of AGN, as it consistently outperformed GN across all tested parameters.  
  
Table 1

AGN vs GN

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Method | Batch size | Group size | Max test accuracy | Last test accuracy |
| GN | 8 | 32 | 64.73 | 64.41 |
| AGN | 8 | 32 | **67.82** | **67.56** |
|  |  |  |  |  |

5. References

[1] Yunxin Wu, Kaiming He. Facebook AI Research. Group Normalization [URL](https://arxiv.org/pdf/1803.08494.pdf)