Dynamic Group Normalization: Spatio-Temporal Adaptation to Evolving Data Statistics

Supplementary Material

Visual Evidence and Intuition

(*Sub-section 3.1.3 in the main paper body*)

To accurately interpret Figure A (supplementary material) and Figure 2 (main body), one must distinguish between two types of black regions in the visualizations. In our DGN, these black regions function purely as visualization placeholders to maintain a uniform square format. In contrast, in baseline GN with fixed 16-channel groups, black regions represent actual channel suppression resulting from rigidly grouping statistically heterogeneous channels.
Figure A demonstrates three key DGN advantages:

- **Benchmark Robustness:** DGN maintains channel similarity within groups, showing superior performance in long-tailed distributions (Figure Ac) and OOD detection (Figure Ad).
- **Architecture Versatility:** Similarity-based grouping remains effective across models of varying complexity, highlighting DGN's model-agnostic benefits.
- **Task Generalization:** DGN adapts successfully across diverse computer vision tasks from image classification to object detection.

As evidenced in both figures, GN's rigid approach forces fixed-size groups that normalize statistically diverse channels together, causing significant feature degradation visible as flat, textureless channels with near-complete darkening or whitening. DGN's similarity-driven grouping preserves structural patterns and discriminative features, maintaining robust representations.

General Hyper-parameters Calibration

(*Sub-section 4.2.1 in the main paper body)*

As discussed in the paper introduction, both GN and DGN were mainly designed for small-batch analysis (Table A). To ensure a fair comparison, we maintained the same batch sizes across all normalization techniques while optimizing the remaining hyperparameters detailed in section 4.2.1. These parameters were tuned through grid search, starting from standard literature values, to enhance performance under small-batch conditions.
Through grid search optimization, we also established an initial group size of $C_{G_{\text{def}}} = 16$ channels for both GN and DGN, aligned with the optimal value reported in [52]. While GN maintained this fixed configuration throughout training, DGN used it only as a starting point to calculate the total number of groups before dynamically adapting their content based on channels' statistics. Our adaptive approach demonstrates superior stability, with DGN exhibiting only 0.26% maximal performance variation across different $C_{G_{\text{def}}}$ values, compared to GN's more substantial 1.79% variation, highlighting DGN's reduced sensitivity to initial parameter settings.

Table A. Batch size settings for various models and benchmarks. VT - visual transformer-based models, CNN - other convolutional-based models.

Dataset	Model-base	Batch size
CIFAR-100	CNN	16
	VT	8
Cityscapes	CNN	4
	VT	2
ImageNet	CNN	16
	VT	8
COCO	YoloV5	8
	YoloV9	4

Spatio-Temporal Group Adaptation

(Sub-section 5.1 in the main paper body)

Figure B provides additional examples of DGN's spatio-temporal adaptation of group size distributions for different datasets and models. It illustrates how DGN adapts group sizes in response to evolving feature statistics, optimizing representation learning. The results highlight DGN's ability to flexibly adjust groupings across diverse architectures, including CNN-based and Transformer-based ones, reinforcing its generalizability across both classification and segmentation tasks.

Analysis of Long-Tail Special Scenario

(Sub-section 5.5.1 in the main paper body)

Table B presents a detailed comparison of GN and DGN under long-tailed data distributions. The results decisively confirm DGN's superior adaptability to various class-imbalanced settings, illustrating its role as a highly effective normalization strategy for real-world applications. Its robust performance across architectures, class distributions, and learning paradigms highlights its potential for enhancing model's performance in long-tailed scenarios:

Table B. **Long-tailed ImageNet-LT**. Superior results for each experiment are highlighted in **bold**. Many (classes with >100 images), Medium (classes with 20-100 images), and Few (classes with <20 images). Regular font size numbers are the mean values over 3 weights initializations, subscript - std values. All models were trained from **scratch** and were validated against literature.

Architecture	Norm	Many	Medium	Few	All
MobilenetV3-small	GN	$56.06_{0.51}$	$32.53_{0.34}$	$10.58_{0.34}$	$38.20_{0.19}$
	DGN	$\mathbf{58.62_{0.44}}$	$\mathbf{34.39_{0.16}}$	$\mathbf{11.81_{0.33}}$	$\mathbf{40.58_{0.41}}$
EfficientNet-B3	GN	$59.31_{0.56}$	$33.41_{0.47}$	$10.13_{0.30}$	$40.87_{0.51}$
	DGN	$\mathbf{61.81_{0.48}}$	$\mathbf{36.23_{0.43}}$	$\mathbf{12.22_{0.40}}$	$\mathbf{43.43_{0.47}}$
ResNet101 (BALLAD)	GN	$72.21_{0.33}$	$64.68_{0.53}$	$59.13_{0.48}$	$67.17_{0.39}$
	DGN	$\mathbf{73.25_{0.47}}$	$\mathbf{66.61_{0.41}}$	$\mathbf{60.77_{0.62}}$	$\mathbf{68.41_{0.47}}$
ViT-Base (BALLAD)	GN	$75.65_{0.26}$	$69.81_{0.47}$	$65.44_{0.34}$	$71.22_{0.28}$
	DGN	$\mathbf{77.28_{0.25}}$	$\mathbf{71.98_{0.31}}$	$\mathbf{67.11_{0.51}}$	$\mathbf{73.14_{0.30}}$
ViT-Base (LIFT)	GN	$76.09_{0.39}$	$72.14_{0.33}$	$67.40_{0.40}$	$73.67_{0.32}$
	DGN	$\mathbf{78.67_{0.22}}$	$\mathbf{74.91_{0.17}}$	$\mathbf{70.02_{0.32}}$	$\mathbf{76.33_{0.26}}$

- **Model-Agnostic Benefits:** DGN demonstrates strong scalability across both lightweight models such as MobileNetV3 and EfficientNet-B3 and deeper state-of-the-art Transformer-based architectures, achieving average accuracy gains of 2.06% and 2.29%, respectively. Its effectiveness is evident regardless of model depth or parameter efficiency. Furthermore, it enhances cutting-edge long-tailed learning techniques such as BALLAD and LIFT, proving its key role for class-imbalanced learning solutions.
- **Class Distribution Analysis:** DGN demonstrates robust performance improvements across all class frequency categories, from well-represented (Many) to extremely underrepresented (Few) classes, ensuring robust learning across the dataset. DGN achieves an average accuracy improvement of 2.06% across multiple architectures in many-shot learning scenarios. Notably, it sustains high accuracy gains even in relatively high accuracy that exceeds 75%. This demonstrates that DGN is not merely beneficial for underrepresented classes but enhances performance holistically, ensuring improvements across the entire data distribution. In medium-shot settings, DGN consistently outperforms GN, with an average accuracy boost of 2.25%. Its strongest impact is observed in state-of-the-art LIFT models, where it achieves a 2.77% gain, highlighting its synergy with advanced class-imbalance strategies. Additionally, DGN exhibits reduced variance compared to GN, leading to greater model stability and more reliable learning dynamics in medium-shot scenarios. For few-shot learning, DGN maintains a 2.04% performance gain even under severe class imbalance, where only a minimal number of instances represent certain classes. Despite the extreme data scarcity, DGN's relative improvement remains consistent with its benefits in many- and medium-shot settings.

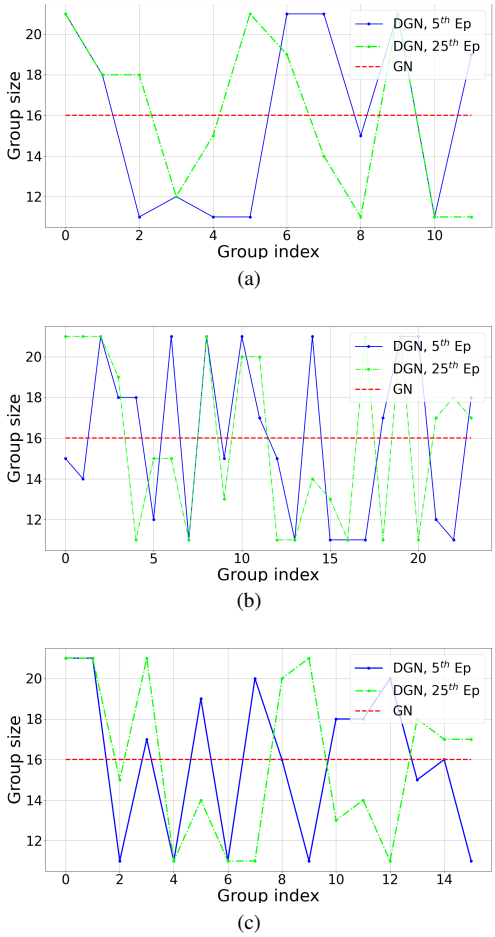

Figure B. Spatio-Temporal group sizes adaptation. The red line indicates GN with fixed group sizes, while the blue and green lines represent DGN at different training epochs, demonstrating the spatial and temporal group sizes changes during different regrouping steps within the same layer. (a) CvT-13 on ImageNet, (b) MobileNetV3 on CIFAR-100, (c) SegFormer-MiT-B0 on Cityscapes.

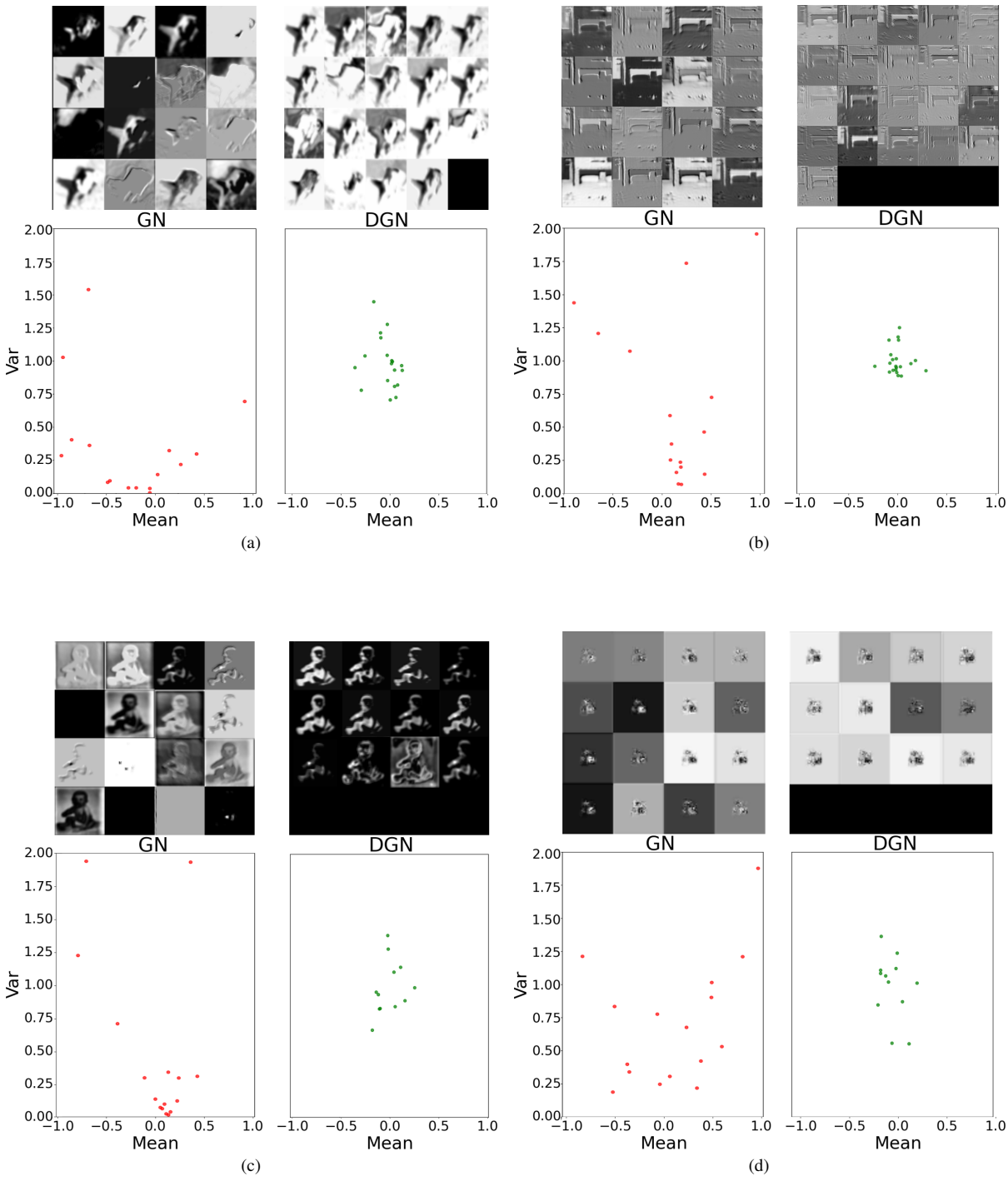

Figure A. Channel Groups Representation in DGN and GN. For each normalization (DGN and GN), a randomly selected group of channels alongside its corresponding scatter plot are presented. Each point in the plot represents an individual channel's statistics (mean, variance). Results are shown across diverse architectures and datasets: (a) RevViT on CIFAR-100, (b) CvT-13 on ImageNet, (c) ResNet-101 (BALLAD) on ImageNet-LT, and (d) YOLOv5L trained on PASCAL-VOC (In-Distribution) and evaluated on MS-COCO (Out-of-Distribution).