# Supplementary Material
## MGiaD: Multigrid in all dimensions.
## Efficiency and robustness by weight sharing and coarsening in resolution and channel dimension

Anonymous ICCV submission

Paper ID 18

## Supplemental Experiments

The following supplemental experiments disseminate some aspects of our multigrid in all dimensions architecture in more detail. These include experiments with grouped convolutions in MgNet and investigations of the influence of a channel scaling parameter $\lambda$ and post-smoothing in MGiaD, on both CIFAR-10 and CIFAR-100. The experimental setup is identical to the main paper.

### 0.1. Evaluation on CIFAR-10

**Grouped Convolutions in MgNet** To justify the use of a hierarchical structure in the channel dimension, we discuss replacing fully coupled convolutions by grouped convolutions in the ResNet-type architecture MgNet. The reduced number of connections between the grouped blocks reduces the weight count and thus the channel interaction. In table 1 the influence of the group size $g_s$ on the number of weights and the accuracy is reported. Compared to ResNet18 and MgNet, both with fully coupled convolutions, smaller group sizes $g_s$ in $A$ and $B$ in MgNet convolutions lead to a drastic reduction of the weight count, but also of the accuracy. A model with convolutions of a rather large group size $g_s = 32$ has $425k$ parameters but only achieves an accuracy of 92 pp. The highest accuracy of MgNet with grouped $A$ and $B$ achieved with a $g_s = 64$ with $831k$ weights. These results suggest a naive sparsification of the channel dimension without a hierarchical structure is not easily achieved.

| Model | $g_s$ | # weights ($k$) | accuracy (%) ± std | |
|---|---|---|---|---|
| | | | test | train |
| ResNet18 | $c$ | 11,170 | **96.26** (0.16) | 98.14 |
| MgNet$^{A,B}$ | $c$ | 2,751 | 96.00 (0.27) | 97.60 |
| | 64 | 831 | 93.88 (0.15) | 95.50 |
| | 32 | 425 | 92.05 (0.28) | 92.86 |
| | 16 | 223 | 89.74 (0.59) | 88.08 |
| | 8 | 121 | 87.02 (0.76) | 82.32 |

Table 1: Results of MgNet with different group sizes $g_s$ in $A$ and $B$ convolutions in a 4 resolution layer MgNet as well as ResNet18 as a reference point on CIFAR-10. The group size for a fully coupled convolution equals to the number of channels $c$. In each table section the model with highest overall accuracy is highlighted with purple background.

| Model | $g_s$ | $\lambda$ | $c_K$ | #weights ($k$) | accuracy (%) ± std | |
|---|---|---|---|---|---|---|
| | | | | | test | train |
| ResNet20 | - | - | - | 270 | **92.44** (0.16) | 93.54 |
| MgNet$^{A,B}$ | - | - | - | 101 | 91.26 (0.09) | 90.88 |
| ResNet18 | - | - | - | 11,170 | **96.26** (0.16) | 98.14 |
| MgNet$^{A,B}$ | - | - | - | 2,751 | 96.00 (0.27) | 97.60 |
| MGiaD | 4 | 1 | 4 | 138 | 90.63 (0.16) | 88.72 |
| | | | 8 | 139 | 90.82 (0.14) | 88.93 |
| | | | 16 | 148 | 91.81 (0.35) | 90.41 |
| | | | 32 | 193 | 93.01 (0.22) | 93.13 |
| | | | 64 | 393 | **94.62** (0.10) | 95.83 |
| | 8 | 1 | 8 | 236 | 92.48 (0.37) | 92.46 |
| | | | 16 | 240 | 92.62 (0.16) | 92.69 |
| | | | 32 | 276 | 93.66 (0.25) | 94.29 |
| | | | 64 | 458 | **94.58** (0.25) | 96.11 |
| | 16 | 1 | 16 | 424 | 94.13 (0.15) | 95.29 |
| | | | 32 | 441 | 94.31 (0.14) | 95.70 |
| | | | 64 | 586 | **94.72** (0.35) | 96.44 |
| | 32 | 1 | 32 | 773 | 94.97 (0.37) | 96.65 |
| | | | 64 | 845 | **95.20** (0.20) | 96.81 |
| | 64 | 1 | 64 | 1,361 | **95.47** (0.07) | 97.23 |

Table 2: Influence of size of coarsest level $c_K$ and group size $g_s$ on accuracy and weights for MGiaD on CIFAR-10.

**Fully Coupled Channels and Group Size in MGiaD** To identify the parameters that play a role in the accuracy-weight trade-off, we performe a large-scale parameter study on CIFAR-10. We start by studying results with varying group sizes $g_s$ and size of coarsest level $c_K$, i.e. fully coupled convolutions, within the channel MG subcycle. In table 2 we study the correlations of the $c_K$ and $g_s$ combinatorially. The results clearly indicate that the size of the fully coupled channels on the coarsest level $\kappa$ of the MG subcycle has a significant influence on the performance of the resulting network, while a small group size affects the weight count strongly. An MGiaD model with the same

ICCV #18

| Model | $g_s$ | $\lambda$ | #weights ($k$) | accuracy(%) ± std | |
|---|---|---|---|---|---|
| | | | | test | train |
| ResNet18 | $c$ | - | 11,174 | **96.26** (0.16) | 98.14 |
| MgNet[A,B] | $c$ | - | 2,751 | 96.00 (0.27) | 97.60 |
| MGiaD | 4 | 1 | 393 | 94.62 (0.10) | 95.83 |
| | | 2 | 533 | 95.02 (0.21) | 96.32 |
| | | 3 | 1,020 | **95.64** (0.09) | 97.21 |
| | 8 | 1 | 458 | 94.58 (0.25) | 96.11 |
| | | 2 | 533 | 95.02 (0.21) | 96.32 |
| | | 3 | 1,269 | **95.95** (0.12) | 97.44 |

Table 3: Number of initial channels multiplied by $\lambda \in \{1, 2, 3\}$, with $g_K = 64$ and $g_s = 4$ and $g_s = 8$ in MGiaD. ResNet and MgNet results for comparison.

weight count as ResNet20, increases the accuracy by 1 pp. With a group size of 8 and $c_K = 64$ the weight count of MGiaD is reduced by a factor of 24 with a small drop in accuracy by 1.5 pp, compared to ResNet18. These findings show, that our MGiaD is very efficient with fewer weights than its residual counterpart.

**Channel Scaling in MGiaD** In previous studies we showed, that a group size of 4 or 8 with $c_K = 64$ fully connected channels on the coarsest in channel level reduces the accuracy while maintaining accuracy. To utilize the freed-up capacity in terms of weight we introduce a channel scaling parameter $\lambda$, which is multiplied to the initial number of channels. We report results for $\lambda \in \{1, 2, 3\}$ in table 3. Multiplying the number of channels naturally increases the the number of weights linearly, which has beneficial effect on the accuracy. For a group size of 8 and $\lambda = 3$ the resulting model achieves almost 96 pp in accuracy, with $1{,}269k$ parameters, which is half the weight count of MgNet and a factor 8 of the ResNet18 parameters. This shows that, as long as the overall cost w.r.t. the channel dimension is linear, as in MGiaD, trading channel connectivity for channel dimension is beneficial.

**Post-smoothing in MGiaD** From the multigrid perspective taken in the main document it is obvious to ask for a variation of the number of smoothing iterations performed on each level of the channel hierarchy. Due to weight sharing of the convolutions an increase in smoothing iteration amount to a marginal increase in parameter. Thus as a last study, we report results of a slightly more complex MGiaD model with $g_s = 8$ and $g_s = 4$ and $\lambda = 3$ for varying number of post-smoothings $\eta_{\text{post}}$ in table 4. Consistently to observations made in table 3, table 2 and fig (7) in the main paper, a bigger group size improves the accuracy, while the effect of number of post-smoothing iterations is ambiguous.

| Model | $g_s$ | $\eta_{\text{post}}$ | #weights ($k$) | accuracy (%) ± std | |
|---|---|---|---|---|---|
| | | | | test | train |
| MGiaD | 4 | 1 | 1,020 | 95.64 (0.09) | 97.21 |
| | | 2 | 1,035 | **95.83** (0.13) | 97.25 |
| | | 3 | 1,051 | 95.53 (0.17) | 97.32 |
| | 8 | 1 | 1,269 | **95.95** (0.12) | 97.44 |
| | | 2 | 1,276 | 95.70 (0.26) | 97.35 |
| | | 3 | 1,300 | 95.90 (0.23) | 97.31 |

Table 4: Influence of post-smoothing $\eta_{\text{post}}$ in MGiaD with channel scale of $\lambda = 3$, $c_K = 64$ and $g_s \in \{4, 8\}$ on CIFAR-10.

| Model | $\lambda$ | $\eta_{\text{post}}$ | $g_s$ | #weights ($k$) | accuracy ± std | |
|---|---|---|---|---|---|---|
| | | | | | test | train |
| ResNet18 | - | - | - | 11,220 | **75.42** (0.13) | 99.98 |
| MgNet | - | - | - | 2,774 | 74.42 (0.28) | 99.98 |
| MGiaD | 1 | 1 | 8 | 481 | 69.91 (0.37) | 99.25 |
| | | | 64 | 1,384 | **72.53** (0.45) | 99.97 |
| | 1 | 3 | 8 | 490 | 70.19 (0.10) | 99.35 |
| | | | 64 | 1,393 | **72.74** (0.73) | 99.96 |
| | 3 | 1 | 8 | 1,338 | 72.75 (0.62) | 99.97 |
| | | | 64 | 4,822 | **75.85** (0.14) | 99.98 |
| | | 3 | 8 | 1,338 | 72.75 (0.38) | 99.96 |
| | | | 64 | 4,853 | **75.40** (0.35) | 99.97 |

Table 5: Influence of number of channels w.r.t. resolution levels, scaled by $\lambda$ and post-smoothing $\eta_{\text{post}}$ for MGiaD on CIFAR-100. The number of fully coupled channels is $c_K = 64$ and the group size is $g_s \in \{8, 64\}$. ResNet18 and corresponding MgNet results for comparison.

## 0.2. Evaluation on CIFAR-100

Similar to CIFAR-10 we study the influence of the number of channels scaled by $\lambda$ and the number of post-smoothings w.r.t. the channel levels. The results are reported in table 5. Consistently to observations made for CIFAR-10, table 3 and table 4, increasing the channel dimension is beneficial for the accuarcy, while the impact of post-smoothing has a negligible effect.