Supplementary for Towards Regression-Free Neural Networks for Diverse Compute Platforms

Rahul Duggal* Hao Zhou Shuo Yang
Jun Fang Yuanjun Xiong Wei Xia

AWS/Amazon AI
rduggal7@gatech.edu {zhouho, shuoy, junfa, yuanjx, wxia}@amazon.com

A Implicit reduction of negative flips through weight sharing

We hypothesize that weight sharing leads to lower negative flips. Fig. 1 presents empirical results to support this hypothesis.

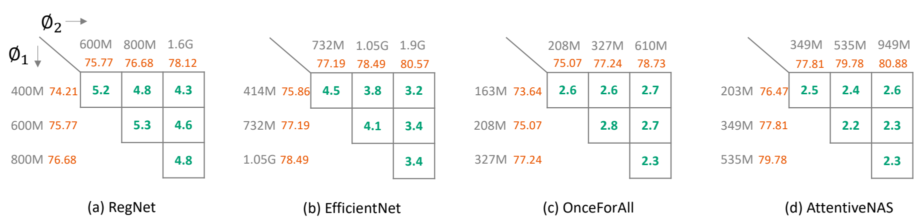


Fig. 1: We present the Top-1 accuracy (in orange), pair-wise negative flip rate (in green) and flops (in black) for a family of neural networks obtained via four model design algorithms. The algorithms in (a) and (b) re-train each model independently and lead to much higher NFR compared to One-Shot NAS algorithms in (c) and (d) that jointly train all sub-networks in a super-network.

Fig. 1 plots the negative flip rate and Top-1 accuracies from four popular model families—RegNet [2], EfficientNet [3], Onceforall [1] and AttentiveNAS [4]. Among these the first two ie. RegNet and EfficientNet independently train the models of different sizes while the latter two i.e. Onceforall and AttentiveNAS sample networks from a common super-network and benefit from weight sharing. From the figure, we can clearly see that the for similar gaps in accuracies, the latter two methods which benefit from weight-sharing (i.e. Fig. 1c,d) lead to much lower negative flip rate than the former two (i.e. Fig. 1a,b).

* Currently at Georgia Tech. Work conducted during an internship with AWS AI.

B Exploring the Top-1 accuracy *v.s.* NFR tradeoff

$\frac{\lambda_2}{\lambda_1}$	Mflops	Top-1 Accuracy $\uparrow$(%)	NFR $\downarrow$(%)
0.05	297	76.91	2.67
0.1	295	77.09	2.70
0.2	298	77.03	2.63
0.5	299	76.94	2.27
1	295	76.86	2.27
2	294	76.94	2.24
5	290	76.74	2.15
10	295	76.78	2.19
20	270	76.15	2.08

Table 1: We measure the impact of varying the weighting factors λ_1 and λ_2 on the Top-1 accuracy and NFR of the searched models. Given a fixed 150 Mflop reference model from the MobileNet-V3 search space of OFA, we search a 300 Mflops target model while varying λ_1 and λ_2. Observe that large values of λ_1 prioritize Top-1 accuracy leading to high Top-1, at the cost of high NFR. Large values of λ_2 prioritizes NFR leading to low NFRs at the loss of low accuracy. Overall, setting $\lambda_1 = \lambda_2 = 1$ leads to reasonable trade-off for both the Top-1 accuracy and NFR.

The search reward of Eq. 4 equally weighs the NFR and Top-1 accuracy by setting $\lambda_1 = \lambda_2 = 1$. In Table 1, we explore the NFR *v.s.* Top-1 trade-off by setting different values of the λ multipliers *i.e.* $\frac{\lambda_2}{\lambda_1} \in [0.05, 20]$. The results follow the expected trend wherein for $\frac{\lambda_2}{\lambda_1} \in [0.05, 1)$ the Top-1 accuracy is prioritized over NFR which becomes as high as 2.70. At the other end, for $\frac{\lambda_2}{\lambda_1} \in (1, 20]$ the NFR is prioritized over Top-1 accuracy so that NFR becomes as low as 2.08. Overall, equally weighing the two metrics, *i.e.* $\frac{\lambda_2}{\lambda_1} = 1$ leads to a resonably high Top-1 accuracy with a reasonably low NFR. Thus we use $\lambda_1 = \lambda_2 = 1$ for REG-NAS.

C How significant are the results of REG-NAS?

We strive to find models that achieve a high Top-1 accuracy and low negative flip rate (NFR). To demonstrate the significance of our results, we extend Table 1 of the main paper by calculating the *relative change* of Top-1 accuracy and NFR which is defined as below:

$$\mathcal{C}(\phi_1, \phi_2) = \frac{\mathcal{M}(\phi_1) - \mathcal{M}(\phi_2)}{\mathcal{M}(\phi_2)}, \tag{1}$$

where $\mathcal{M} \in \{\text{Top-1}, \text{NFR}\}$ and ϕ_1, ϕ_2 are two models.

	Model	Top-1↑(%) Abs.	Top-1↑(%) $\mathcal{C}$	NFR↓ (%) Abs.	NFR↓ (%) $\mathcal{C}$
ref	MB-$\mathcal{R}_0$-150 [1]	73.70		-	
baseline	EffNet	77.08	-0.30	4.25	-49.1
PCT	EffNet+FD [5]	76.25	+0.76	3.25	-33.5
wt. share	MB-$\mathcal{R}_0$-300 [1]	77.11	-0.37	2.51	-13.9
proposed	MB-($\mathcal{R}_2$+CAS)-300	76.83	0	2.16	0

(a) MobileNet-V3 search space of OFA

	Model	Top-1↑(%) Abs.	Top-1↑(%) $\mathcal{C}$	NFR↓ (%) Abs.	NFR↓ (%) $\mathcal{C}$
ref	RN-$\mathcal{R}_0$-2000 [1]	78.25		-	
baseline	RN101	79.21	-0.51	4.83	-67.4
PCT	RN101+FD [5]	79.90	-1.37	3.06	-48.6
wt. share	RN-$\mathcal{R}_0$-3000 [1]	78.78	-0.02	2.04	-23.0
proposed	RN-($\mathcal{R}_2$+CAS)-3000	78.80	0	1.57	0

(b) ResNet search space of OFA

Table 2: Extending Table 1 from the main paper. In addition to the absolute values (abs.), we present the relative change ($\mathcal{C}$) of Top-1 accuracy and NFR. EffNet and RN101 represents EfficientNet-B0 and ResNet-101 trained with cross entropy loss while EffNet+FD and RN101+FD are trained with state-of-the-art focal distillation loss [5]. Results of weight sharing and our proposed method are averaged from three runs with different random seeds. Observe that REG-NAS leads to a large relative change in NFR at a marginal expense of the Top-1 accuracy.

Table 2 presents the relative change of Top-1 accuracy and NFR between the model searched via REG-NAS (considered as ϕ_1) and other models (considered as ϕ_2). The results show that REG-NAS leads to large relative reduction of NFR *e.g.* upto 50% and 33.5% w.r.t. baseline and state-of-the-art on MobileNet-V3 super-network and upto 67.4% and 48.6% w.r.t. baseline and state-of-the-art on the ResNet-50 super-network of OFA[1]. This reduction in NFR is despite very little (*e.g.* less than 1%) relative change of Top-1 accuracy. These results clearly demonstrate that REG-NAS can significantly reduce NFR at minimal expense of Top-1 accuracy.

D Extending the search to 4 architectures

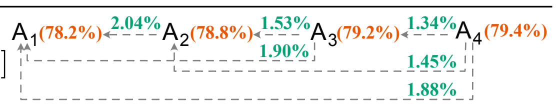
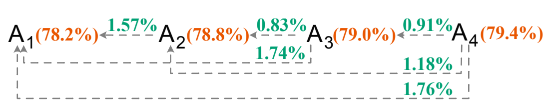
Method	Results
OFA [1]	
REG-NAS	

Table 3: Testing generalization of REG-NAS for searching 4 models with diverse compute budgets from the ResNet search space of OFA [1]. Model size increases from A_1 to A_4. NFR is indicated in green and Top-1 accuracy in orange. REG-NAS successfully reduces NFR in all scenarios.

Table. 4 of the main paper presents the NFR and Top-1 accuracy for three models *i.e.* A1 (2000 Mflops), A2 (3000 Mflops), A3 (4000 Mflops) from the ResNet search space of OFA [1]. However, one could also extend the search to a larger family of models. Table 3 extends the search to a larger model A4 (5000 Mflops) from the ResNet search space of OFA [1]. Observe that the four models A1-A4 searched via REG-NAS achieve significantly smaller pairwise NFR in all cases, with minimal loss of Top-1 accuracy. This demonstrates that REG-NAS generalizes to searching for regression-free models across multiple flops budgets. Note that for this experiment, we only consider the ResNet space of OFA since the MobileNet-V3 space cannot search sub-networks larger than A3 (600 Mflops).

References

1. Cai, H., Gan, C., Wang, T., Zhang, Z., Han, S.: Once for all: Train one network and specialize it for efficient deployment. In: ICLR (2020)
2. Radosavovic, I., Kosaraju, R.P., Girshick, R., He, K., Dollár, P.: Designing network design spaces. In: CVPR (2020)
3. Tan, M., Le, Q.: EfficientNet: Rethinking model scaling for convolutional neural networks. In: ICML (2019)
4. Wang, D., Li, M., Gong, C., Chandra, V.: Attentivenas: Improving neural architecture search via attentive sampling. In: CVPR (2021)
5. Yan, S., Xiong, Y., Kundu, K., Yang, S., Deng, S., Wang, M., Xia, W., Soatto, S.: Positive-congruent training: Towards regression-free model updates. In: CVPR (2021)