The Supplementary Material for "Accelerate CNNs from Three Dimensions: A Comprehensive Pruning Framework"

A. Why the Rank of Θ Is 1

A.1. Experiments

In our original paper, we propose a model accuracy predictor (MAP):

$$\mathcal{F}(d,w,r;\Theta)=\sum_{q=1}^{\mathcal{R}}\mathcal{H}(d;\vec{s_q})\mathcal{H}(w;\vec{u_q})\mathcal{H}(r;\vec{v_q}), \tag{1}$$

where $\mathcal{H}$ represents a univariate polynomial, and $\mathcal{R}$ means the rank of Θ. Pruning experiments shown in our original paper are done with $\mathcal{R}=1$ because we find that $\mathcal{R}=1$ is enough for an accurate approximation of MAP. In this case, Eq. (1) becomes:

$$\mathcal{F}(d,w,r;\Theta)=\mathcal{H}(d;\vec{s})\mathcal{H}(w;\vec{u})\mathcal{H}(r;\vec{v}). \tag{2}$$

We assume that $\mathcal{R}=1$ works well because it accords with the real distribution of $\mathcal{F}(d,w,r)$. To further verify the assumption. we design some experiments to show the relations between the model's accuracy and (d,w,r). Specifically, we train many ResNets and DenseNets with different (d,w,r), and the results are in Table 1 — part of them have also been reported in the original paper. The observations are shown in their subtitles, from which we can draw the same conclusion as Eq. (2).

A.2. Proof

Omitting all subscripts of d_1, w_1, and r_1 of the equations in Table 1, we have:

$$\begin{aligned}&\frac{\mathcal{F}(d_2,w,r)}{\mathcal{F}(d_2,w_2,r)}\approx\frac{\mathcal{F}(d_2,w,r_2)}{\mathcal{F}(d_2,w_2,r_2)}\\ \Rightarrow\ &\mathcal{F}(d_2,w,r)\approx\frac{\mathcal{F}(d_2,w_2,r)\mathcal{F}(d_2,w,r_2)}{\mathcal{F}(d_2,w_2,r_2)},\end{aligned} \tag{3}$$

$$\begin{aligned}&\frac{\mathcal{F}(d,w,r)}{\mathcal{F}(d_2,w,r)}\approx\frac{\mathcal{F}(d,w,r_2)}{\mathcal{F}(d_2,w,r_2)}\\ \Rightarrow\ &\mathcal{F}(d,w,r)\approx\frac{\mathcal{F}(d_2,w,r)\mathcal{F}(d,w,r_2)}{\mathcal{F}(d_2,w,r_2)},\end{aligned} \tag{4}$$

$$\begin{aligned}&\frac{\mathcal{F}(d,w,r_2)}{\mathcal{F}(d_2,w,r_2)}=\frac{\mathcal{F}(d,w_2,r_2)}{\mathcal{F}(d_2,w_2,r_2)}\\ \Rightarrow\ &\mathcal{F}(d,w,r_2)=\frac{\mathcal{F}(d_2,w,r_2)\mathcal{F}(d,w_2,r_2)}{\mathcal{F}(d_2,w_2,r_2)}.\end{aligned} \tag{5}$$

Substituting Eq. (3) and Eq. (5) into Eq. (4), we have:

$$\begin{aligned}\mathcal{F}(d,w,r)&\approx\frac{\mathcal{F}(d,w_2,r_2)\mathcal{F}(d_2,w,r_2)\mathcal{F}(d_2,w_2,r)}{\mathcal{F}(d_2,w_2,r_2)^2}\\ &=\frac{\mathcal{H}(d;\vec{s})\mathcal{H}(w;\vec{u})\mathcal{H}(r;\vec{v})}{\mathcal{F}(d_2,w_2,r_2)^2},\end{aligned} \tag{6}$$

Table 1. Accuracies (%) of ResNets and DenseNets on CIFAR-10 with different depths (d), widths (w) and resolutions (r). The base models are ResNet-32 and DenseNet-40.

(a) $\frac{\mathcal{F}(d_2,w_1,r_1)}{\mathcal{F}(d_2,w_2,r_1)}\approx\frac{\mathcal{F}(d_2,w_1,r_2)}{\mathcal{F}(d_2,w_2,r_2)}$

ResNet ($d=1.0$)					
w \ r	1.00	0.87	0.75	0.62	0.50
0.60	90.59	89.43	88.51	86.59	83.19
0.70	91.39	90.53	89.26	87.38	84.62
0.80	92.19	90.95	89.88	88.38	85.15
0.90	92.54	91.55	90.77	88.60	85.93
1.00	92.84	91.87	91.10	89.38	86.56

DenseNet ($d=1.0$)					
w \ r	1.00	0.87	0.75	0.62	0.50
0.60	90.82	90.48	89.62	88.04	85.63
0.70	91.54	90.98	90.19	88.79	86.75
0.80	91.99	91.57	90.59	90.01	87.67
0.90	92.74	92.08	91.19	90.56	88.12
1.00	93.09	92.25	92.05	90.68	88.38

(b) $\frac{\mathcal{F}(d_1,w_1,r_1)}{\mathcal{F}(d_2,w_1,r_1)}\approx\frac{\mathcal{F}(d_1,w_1,r_2)}{\mathcal{F}(d_2,w_1,r_2)}$

ResNet ($w=1.0$)					
d \ r	1.00	0.87	0.75	0.62	0.50
0.11	86.88	85.91	85.15	83.64	81.68
0.33	92.30	91.12	90.08	88.79	85.87
0.55	92.84	91.87	91.10	89.38	86.56
0.77	93.43	92.40	91.77	89.87	87.17
1.00	93.63	92.63	91.79	90.14	87.48

DenseNet ($w=1.0$)					
d \ r	1.00	0.87	0.75	0.62	0.50
0.20	88.16	87.64	86.77	85.50	83.83
0.40	92.00	91.22	90.32	89.15	86.89
0.60	93.03	92.06	91.85	90.63	88.42
0.80	93.80	93.21	92.78	91.74	89.25
1.00	94.53	93.75	93.69	92.21	89.84

(c) $\frac{\mathcal{F}(d_1,w_1,r_2)}{\mathcal{F}(d_2,w_1,r_2)}\approx\frac{\mathcal{F}(d_1,w_2,r_2)}{\mathcal{F}(d_2,w_2,r_2)}$

ResNet ($r=1.0$)					
w \ d	0.11	0.33	0.55	0.77	1.00
0.60	83.36	89.39	90.91	91.41	92.01
0.70	84.03	90.17	91.43	91.85	92.36
0.80	85.45	91.21	92.16	92.65	92.7
0.90	86.14	91.74	92.60	92.56	93.12
1.00	86.22	92.12	92.88	93.17	93.64

DenseNet ($r=1.0$)					
w \ d	0.20	0.40	0.60	0.80	1.00
0.60	84.51	88.82	89.95	91.40	92.70
0.70	85.43	89.26	90.51	92.35	92.83
0.80	86.92	90.75	91.07	93.01	93.66
0.90	87.88	91.74	91.46	93.14	93.86
1.00	88.16	92.00	93.03	93.80	94.53

where $\frac{1}{\mathcal{F}(d_2,w_2,r_2)^2}$ is a constant and can be merged into any $\mathcal{H}$. Thus, Eq. (6) can be re-formulated as:

$$\mathcal{F}(d,w,r) \approx \mathcal{H}(d;\vec{s})\mathcal{H}(w;\vec{u})\mathcal{H}(r;\vec{v}), \tag{7}$$

which complies with Eq. (2).