# Theoretical Investigation of Generalization Bound for Residual Networks Supplementary Material


**Anonymous Author(s)**
Affiliation
e-mail


## 1 Proof Details for Technical Lemmas

In this section, we show the details of our main results:

**Lemma 1.1.** *With fixed values for $t$, function space $\mathcal{F}$ and function space $\mathcal{G}$, the following inequality holds:*

$$\begin{aligned}&\mathbb{E}_\epsilon \exp\left(t\big(\sup_{f\in\mathcal{F}}|\sum_{i=1}^n \epsilon_i f(x_i)| + \sup_{g\in\mathcal{G}}|\sum_{i=1}^n \epsilon_i g(x_i)|\big)\right)\\ \leqslant& 2\mathbb{E}_\epsilon \exp\left(t\big(\sup_{f\in\mathcal{F}}\sum_{i=1}^n \epsilon_i f(x_i) + \sup_{g\in\mathcal{G}}|\sum_{i=1}^n \epsilon_i g(x_i)|\big)\right).\end{aligned}$$

**Lemma 1.2.** *With fixed values for $t, p, q$, and the function space $\mathcal{F}$, we have:*

$$\begin{aligned}&\mathbb{E}_\epsilon \exp\left(t \sup_{\substack{|\mathbf{M}\|_{p,q}\leqslant c\\ f\in\mathcal{F}}} \left\|\sum_{i=1}^n \epsilon_i \sigma\circ(\mathbf{M}(\sigma\circ f(x_i),1)^T)\right\|_{p^*}\right)\\ \leqslant& 2\mathbb{E}_\epsilon \exp\left(tc\rho d^{[\frac{1}{p^*}-\frac{1}{q}]_+}\left(\sup_{f\in\mathcal{F}}\left\|\sum_{i=1}^n \epsilon_i\sigma\circ f(x_i)\right\|_{p^*} + \left|\sum_{i=1}^n \epsilon_i\right|\right)\right),\end{aligned}$$

*where $d$ is the column dimension of matrix $\mathbf{M}$.*

*Proof.*

$$\begin{aligned}&\mathbb{E}_\epsilon \exp\left(t \sup_{\substack{\|\mathbf{M}\|_{p,q}\leqslant c\\ f\in\mathcal{F}}} \left\|\sum_{i=1}^n \epsilon_i \sigma\circ(\mathbf{M}(\sigma\circ f(x_i),1)^T)\right\|_{p^*}\right)\\ \leqslant&\mathbb{E}_\epsilon \exp\Bigg(t\\ &\cdot \sup_{\substack{\|\mathbf{a}\|_q\leqslant c\\ \|\mathbf{V}_j\|_p=\mathbf{a}[j]\\ f\in\mathcal{F}}} \left(\sum_{j=1}^d \left|\sum_{i=1}^n \epsilon_i\sigma\circ(\mathbf{V}_j(\sigma\circ f(x_i),1)^T)\right|^{p^*}\right)^{\frac{1}{p^*}}\Bigg)\\ \leqslant&\mathbb{E}_\epsilon \exp\Bigg(tcd^{[\frac{1}{p^*}-\frac{1}{q}]_+}\\ &\cdot \sup_{\substack{\|\mathbf{V}_j\|_p=1\\ f\in\mathcal{F}}} \left|\sum_{i=1}^n \epsilon_i\sigma\circ(\mathbf{V}_j(\sigma\circ f(x_i),1)^T)\right|\Bigg) \qquad (1)\\ \leqslant& 2\mathbb{E}_\epsilon \exp\left(tc\rho d^{[\frac{1}{p^*}-\frac{1}{q}]_+}\left(\sup_{f\in\mathcal{F}}\left\|\sum_{i=1}^n \epsilon_i\sigma\circ f(x_i)\right\|_{p^*} + \left|\sum_{i=1}^n \epsilon_i\right|\right)\right), \qquad (2)\end{aligned}$$

where step (1) follows Lemma 1.4, step (2) follows the triangle inequality.

□

**Lemma 1.3.** $\forall \mathbf{x}_i\in\mathcal{S}\subset\mathcal{X}; \forall\mathcal{RN}^{k,\mathbf{d}}_{p,q,\mathbf{c}}$;

$$\begin{aligned}&Z_0 \triangleq \left\|\sum_{i=1}^n \epsilon_i\mathbf{x}_i\right\|_{p^*},\\ &Z_j \triangleq \sup_{f\in\mathcal{RN}^{k,\mathbf{d}}_{p,q,\mathbf{c}}} \left\|\sum_{i=1}^n \epsilon_i\sigma\circ f_j(\mathbf{x}_i)\right\|_{p^*},\ 1\leqslant j\leqslant k,\end{aligned}$$

*where $\sigma\circ f_i \triangleq \widetilde{\mathcal{F}}_i\circ\ldots\circ\widetilde{\mathcal{F}}_1$. For any $j=0,1,\cdots,k$ and any $t\in\mathbb{R}$, $\epsilon_i$, the Rademacher Random Variable is:*

$$\begin{aligned}&\mathbb{E}_\epsilon\exp(tZ_j)\\ \leqslant& 8^j\exp\left(\frac{t^2ns_j^2}{2} + \prod_{l=1}^j\left(c_{j,2}c_{j,1}\rho^2(d_{j,1}d_{j,2})^{[\frac{1}{p^*}-\frac{1}{q}]_+}\right)d_0^{\frac{1}{p^*}}\sqrt{n(C(p))}\right),\\ &s_0=1,\\ &s_j=(c_{j,2}c_{j,1}\rho^2(d_{j,1}d_{j,2})^{[\frac{1}{p^*}-\frac{1}{q}]_+}+1)s_{j-1}\\ &\quad + c_{j,2}c_{j,1}\rho^2(d_{j,1}d_{j,2})^{[\frac{1}{p^*}-\frac{1}{q}]_+} + c_{j,2}\rho d_{j,2}^{[\frac{1}{p^*}-\frac{1}{q}]_+},\\ &j=1,2,\cdots,k,\end{aligned}$$

$$C(p)\triangleq\begin{cases}2\log(2d_0) & p\in\{1\}\cup(2,\infty),\\ \min(p^*-1, 2\log(2d_0)) & p\in(1,2].\end{cases}$$

***Proof of Lemma 1.3***. We prove this lemma by induction. When $j=0$, according to Lemma 5 of [Xu and Wang(2018)], $\mathbb{E}_\epsilon Z_0\leqslant d_0^{\frac{1}{p^*}}\sqrt{nC(p)}$. Note that $Z_0$ is a deterministic function of the i.i.d.random variables $\epsilon_1,\epsilon_2,\cdots,\epsilon_n$, which satisfies:

$$\begin{aligned}&|Z_0(\epsilon_1,\cdots,\epsilon_i,\cdots,\epsilon_n) - Z_0(\epsilon_1,\cdots,-\epsilon_i,\cdots,\epsilon_n)|\\ \leqslant& 2\max\|\mathbf{x}_i\|_{p^*}\leqslant 2d_0^{\frac{1}{p^*}}\max\|\mathbf{x}_i\|_\infty\end{aligned}$$

By assuming that $\|\mathbf{x}\|_\infty\leqslant 1$ for any input $\mathbf{x}$ and [Bousquet et al.(2003)Bousquet, Boucheron, and Lugosi], we have:

$$\begin{aligned}\mathbb{E}_\epsilon\exp(tZ_0) =& \mathbb{E}_\epsilon\exp(t(Z_0-\mathbb{E}_\epsilon Z_0)) * \exp(t\mathbb{E}_\epsilon Z_0)\\ \leqslant& \exp\left(\frac{t^2n}{2} + t\sqrt{n(C(p))}\right)\end{aligned}$$

for any $t \in \mathbb{R}$. Then, for $j > 0$,

$$
\begin{aligned}
&\mathbb{E}_\epsilon \exp(tZ_j) \\
=&\mathbb{E}_\epsilon \exp\Bigg(t \sup_{\substack{\|\mathbf{M}_{j,2}\|_{p,q} \leqslant c_{j,2} \\ \|\mathbf{M}_{j,1}\|_{p,q} \leqslant c_{j,1} \\ f \in \mathcal{RN}^{k,\mathbf{d}}_{p,q,\mathbf{c}}}} \Bigg\| f_{j-1}(\mathbf{x}_i) \\
&+ \sum_{i=1}^n \epsilon_i \sigma\big(\mathbf{M}_{j,2}(\sigma \circ \mathbf{M}_{j,1}(\sigma \circ f_{j-1}(\mathbf{x}_i), 1)^T, 1)\big)\Bigg\|_{p^*}\Bigg) \\
\leqslant&\mathbb{E}_\epsilon \exp\Bigg(t \sup_{\substack{\|\mathbf{M}_{j,2}\|_{p,q} \leqslant c_{j,2} \\ \|\mathbf{M}_{j,1}\|_{p,q} \leqslant c_{j,1} \\ f \in \mathcal{RN}^{k,\mathbf{d}}_{p,q,\mathbf{c}}}} \Bigg\{\Bigg\| \sum_{i=1}^n \epsilon_i \sigma \circ f_{j-1}(\mathbf{x}_i)\Bigg\|_{p^*} \\
&+\Bigg\| \sum_{i=1}^n \epsilon_i \sigma\big(\mathbf{M}_{j,2}(\sigma \circ \mathbf{M}_{j,1}(\sigma \circ f_{j-1}(\mathbf{x}_i), 1)^T, 1)^T\big)\Bigg\|_{p^*}\Bigg\}\Bigg) \qquad (3) \\
\leqslant&2\mathbb{E}_\epsilon \exp\Bigg(t \sup_{\substack{\|\mathbf{M}_{j,1}\|_{p,q} \leqslant c_{j,1} \\ f \in \mathcal{RN}^{k,\mathbf{d}}_{p,q,\mathbf{c}}}} \Bigg\{\Bigg\| \sum_{i=1}^n \epsilon_i \sigma \circ f_{j-1}(\mathbf{x}_i)\Bigg\|_{p^*} \\
&+c_{j,2}\rho d_{j,2}^{[\frac{1}{p^*}-\frac{1}{q}]_+}\Bigg\| \sum_{i=1}^n \epsilon_i (\sigma \circ \mathbf{M}_{j,1}(\sigma \circ f_{j-1}(\mathbf{x}_i), 1)^T, 1)\Bigg\|_{p^*}\Bigg\}\Bigg) \qquad (4) \\
\leqslant&4\mathbb{E}_\epsilon \exp\Bigg(t \sup_{f \in \mathcal{RN}^{k,\mathbf{d}}_{p,q,\mathbf{c}}} \Bigg\{c_{j,2}c_{j,1}\rho^2(d_{j,1}d_{j,2})^{[\frac{1}{p^*}-\frac{1}{q}]_+} \\
&\cdot \Bigg\| \sum_{i=1}^n \epsilon_i (1, \sigma \circ f_{j-1}(\mathbf{x}_i))\Bigg\|_{p^*} + \Bigg\| \sum_{i=1}^n \epsilon_i \sigma \circ f_{j-1}(\mathbf{x}_i)\Bigg\|_{p^*}\Bigg\} \\
&+ tc_{j,2}\rho d_{j,2}^{[\frac{1}{p^*}-\frac{1}{q}]_+}|\sum_{i=1}^n \epsilon_i|\Bigg) \qquad (5) \\
\leqslant&4\mathbb{E}_\epsilon \exp\Bigg(t\big(c_{j,2}c_{j,1}\rho^2(d_{j,1}d_{j,2})^{[\frac{1}{p^*}-\frac{1}{q}]_+} + 1\big) \\
&\cdot \sup_{f \in \mathcal{RN}^{k,\mathbf{d}}_{p,q,\mathbf{c}}} \Bigg\| \sum_{i=1}^n \epsilon_i \sigma \circ f_{j-1}(\mathbf{x}_i)\Bigg\|_{p^*} \\
&+ t(c_{j,2}c_{j,1}\rho^2(d_{j,1}d_{j,2})^{[\frac{1}{p^*}-\frac{1}{q}]_+} + c_{j,2}\rho d_{j,2}^{[\frac{1}{p^*}-\frac{1}{q}]_+})|\sum_{i=1}^n \epsilon_i|\Bigg) \\
\leqslant&4\Bigg[2\mathbb{E}_\epsilon \exp\big(r_j t(c_{j,2}c_{j,1}\rho^2(d_{j,1}d_{j,2})^{[\frac{1}{p^*}-\frac{1}{q}]_+} \\
&+ c_{j,2}\rho d_{j,2}^{[\frac{1}{p^*}-\frac{1}{q}]_+})\sum_{i=1}^n \epsilon_i\big)\Bigg]^{\frac{1}{r_j}} \qquad (6) \\
&\cdot\Bigg[\mathbb{E}_\epsilon \exp\Bigg(r_j^* t\big(c_{j,2}j,1\rho^2(d_{j,1}d_{j,2})^{[\frac{1}{p^*}-\frac{1}{q}]_+} + 1\big) \\
&\cdot \sup_{f \in \mathcal{RN}^{k,\mathbf{d}}_{p,q,\mathbf{c}}} \Bigg\| \sum_{i=1}^n \epsilon_i \sigma \circ f_{j-1}(\mathbf{x}_i)\Bigg\|_{p^*}\Bigg)\Bigg]^{\frac{1}{r_j^*}} \qquad (7) \\
\leqslant&8^j \exp\Bigg(\frac{t^2 n s_j^2}{2} + t\prod_{l=1}^{j}\big(c_{j,2}c_{j,1}\rho^2(d_{j,1}d_{j,2})^{[\frac{1}{p^*}-\frac{1}{q}]_+} + 1\big)\sqrt{nC(p)}\Bigg).
\end{aligned}
$$

Here, step (3) follows the triangle inequality, then step (4) and step (5) follow Lemma 1.2. Step (7) holds for $r_j > 1$ and $\frac{1}{r_j} + \frac{1}{r_j^*} = 1$ according to the Hölder Inequality $\mathbb{E}(|XY|) \leqslant \mathbb{E}(|X|^r)^{\frac{1}{r}} \cdot \mathbb{E}(|Y|^{r^*})^{\frac{1}{r^*}}$. Since $|\sum_{i=1}^n \epsilon_i|$ also satisfies:

$$\big||\sum_{i=1}^n \epsilon_i| - |\epsilon_1 + \cdot + (-\epsilon_i) + \cdot + \epsilon_n|\big| \leqslant 2|\epsilon_i| = 2.$$

According to [Bousquet et al.(2003)Bousquet, Boucheron, and Lugosi], we have the inequation: $\mathbb{E}_\epsilon \exp(t\sum_{i=1}^n \epsilon_i) \leqslant \exp(\frac{t^2 2}{2})$. Thus, by taking

$$r_j = \frac{c_{j,2}c_{j,1}\rho^2(d_{j,1}d_{j,2})^{[\frac{1}{p^*}-\frac{1}{q}]_+} + 1}{c_{j,2}c_{j,1}\rho^2(d_{j,1}d_{j,2})^{[\frac{1}{p^*}-\frac{1}{q}]_+} + c_{j,2}\rho d_{j,2}^{[\frac{1}{p^*}-\frac{1}{q}]_+}} \cdot s_{j-1} + 1$$

and using the induction assumption, we obtain the final result. □

**Proof of Theorem 3.2.**

$$
\begin{aligned}
&n\hat{\mathfrak{R}}_{\mathcal{S}}(\mathcal{RN}^{k,\mathbf{d}}_{p,q,\mathbf{c}}) \\
=&\mathbb{E}_\epsilon\Bigg[\sup_{f \in \mathcal{RN}^{k,\mathbf{d}}_{p,q,\mathbf{c}}} \sum_{i=1}^n \epsilon_i f(\mathbf{x}_i)\Bigg] \\
\leqslant&\frac{1}{t}\log \mathbb{E}_\epsilon \exp\Bigg(t \sup_{f \in \mathcal{RN}^{k,\mathbf{d}}_{p,q,\mathbf{c}}} \sum_{i=1}^n \epsilon_i f(\mathbf{x}_i)\Bigg) \\
\leqslant&\frac{1}{t}\log \mathbb{E}_\epsilon \exp\Bigg(tc_{k+1}\rho d_{k+1}^{[\frac{1}{p^*}-\frac{1}{q}]_+} \\
&\cdot \sup_{f \in \mathcal{RN}^{k,\mathbf{d}}_{p,q,\mathbf{c}}} \Bigg\| \sum_{i=1}^n \epsilon_i (1, \sigma \circ f_k(\mathbf{x}_i))\Bigg\|_{p^*}\Bigg) \\
\leqslant&\frac{1}{t}c_{k+1}\rho d_{k+1}^{[\frac{1}{p^*}-\frac{1}{q}]_+}\Bigg((3k+2)\log 2 + \frac{nt^2 s_k^2}{2} \\
&+ t\prod_{j=1}^{k}\big(c_{j,2}c_{j,1}\rho^2(d_{j,1}d_{j,2})^{[\frac{1}{p^*}-\frac{1}{q}]_+} + 1\big)\sqrt{nC(p)}\Bigg)
\end{aligned}
$$

If we choose $t = \frac{\sqrt{(6k+4)\log 2}}{\sqrt{n}s_{k+1}}$, then

$$
\begin{aligned}
&\hat{\mathfrak{R}}_{\mathcal{S}}(\mathcal{RN}^{k,\mathbf{d}}_{p,q,\mathbf{c}}) \\
\leqslant&c_{k+1}\rho d_{k+1}^{[\frac{1}{p^*}-\frac{1}{q}]_+}\Bigg(\sqrt{\frac{(6k+4)\log 2}{n}}s_{k+1} \\
&+ \prod_{l=1}^{j}\big(c_{j,2}c_{j,1}\rho^2(d_{j,1}d_{j,2})^{[\frac{1}{p^*}-\frac{1}{q}]_+} + 1\big)\sqrt{nC(p)}\Bigg)
\end{aligned}
$$

□

**Lemma 1.4.** *[Xu and Wang(2018)] $\forall p, q \geqslant 1, s_1, s_2 \geqslant 1, \epsilon \in \{1, -1\}^n$, and all functions $h : \mathbb{R}^{m_1} \to \mathbb{R}^{s_1}$, we have:*

$$
\begin{aligned}
&\sup_{\mathbf{M} \in \mathbb{R}^{s_1 \times s_2}} \frac{1}{\|\mathbf{M}\|_{p,q}} \Bigg\| \sum_{i=1}^n \epsilon_i \cdot \sigma\left(\mathbf{M}^T h(\mathbf{x}_i)\right)\Bigg\|_{p^*} \\
\leqslant& s_2^{[\frac{1}{p^*}-\frac{1}{q}]_+} \sup_{\mathbf{v} \in \mathbb{R}^{s_1}} \frac{1}{\|\mathbf{v}\|_p} |\epsilon_i \cdot \sigma[\langle \mathbf{v}, h(\mathbf{x}_i)\rangle]| \,.
\end{aligned}
$$

**Lemma 1.5.** *[Ledoux and Talagrand(2013)] An upper bound of the Rademacher Complexity of $\mathcal{RG}_\gamma$ is:*

$$\hat{\mathfrak{R}}_S(\mathcal{RG}_\gamma) \leqslant \gamma \cdot \hat{\mathfrak{R}}_S(\mathcal{RN}^{k,\mathbf{d}}_{p,q,\mathbf{c}}).$$

**Lemma 1.6.** *[Mohri et al.(2012)Mohri, Talwalkar, and Rostamizadeh] Let $\mathbf{z} \triangleq (\mathbf{x}, y) \sim \mathcal{D}, (\mathbf{x}, y) \in \mathcal{X} \times \mathcal{Y}$, $\mathcal{S} \triangleq \{\mathbf{z}_1, ..., \mathbf{z}_n\}$ is a dataset of m i.i.d samples selected from the distribution $\mathcal{D}$. Let $\mathcal{N} \subset \{f | f : \mathcal{X} \times \mathcal{Y} \to [a, a+1]\}$ Fix $\delta \in (0,1), \forall k \in \mathbb{N}^+, \forall d_i \in \mathbb{N}^+ \quad i = 1, ..., k$. With probability of at least $1-\delta$ over the generation of $\mathcal{S}$, it holds that:*

$$\mathbb{E}_{\mathcal{D}}[g] - \hat{\mathbb{E}}_{\mathcal{S}}[g] \leqslant 2\mathfrak{R}_n(\mathcal{N}) + \sqrt{\frac{\log(1/\delta)}{2n}}.$$

**Proof of Theorem 4.3**

*Proof.* With Lemma 1.6, we have a probability of at least $1-\delta$:

$$\begin{aligned}
\mathbb{E}_{\mathcal{D}}[g] - \hat{\mathbb{E}}_{\mathcal{S}}[g] &\leqslant 2\mathfrak{R}_n(\mathcal{RG}_\gamma) + \sqrt{\frac{\log(1/\delta)}{2n}} \\
\hat{\mathbb{E}}_{\mathcal{S}}[g] - \mathbb{E}_{\mathcal{D}}[g] &= \mathbb{E}_{\mathcal{D}}[-g] - \hat{\mathbb{E}}_{\mathcal{S}}[-g] \\
&\leqslant 2\mathfrak{R}_n(\mathcal{RG}_\gamma) + \sqrt{\frac{\log(1/\delta)}{2n}} \\
\implies \left|\mathbb{E}_{\mathcal{D}}[g] - \hat{\mathbb{E}}_{\mathcal{S}}[g]\right| &\leqslant 2\mathfrak{R}_n(\mathcal{RG}_\gamma) + \sqrt{\frac{\log(1/\delta)}{2n}}.
\end{aligned}$$

According to Lemma 1.5 :

$$\mathfrak{R}_n(\mathcal{RG}_\gamma) \leqslant \sup_{\mathcal{S}} \hat{\mathfrak{R}}_{\mathcal{S}}(\mathcal{RG}_\gamma) \leqslant \gamma \cdot \hat{\mathfrak{R}}_{\mathcal{S}}(\mathcal{RN}^{k,\mathbf{d}}_{p,q,\mathbf{c}}).$$

The combination of the results above and Theorem 3.2 lead to the ultimate conclusion.

□

## 2 Full data for Numerical Experiments

In this section, we display all the data from numerical experiments. seWe train two simple networks and calculate the generalization bound for each network. While both of the networks share the same initialization and parameters, ResNets shortcut structure is added to the second network.

### 2.1 Experiments Design

We first set Net-A and Net-B as two fully connected DNNs with four hidden layers. Then, we add a residual shortcut to Net-B between the second layer and the third one. The parameters are shown in Figure 2. Through the course of several experiments, we found that the choice of widths did not greatly affect the general conclusion; hence, we arbitrarily selected the width parameters.

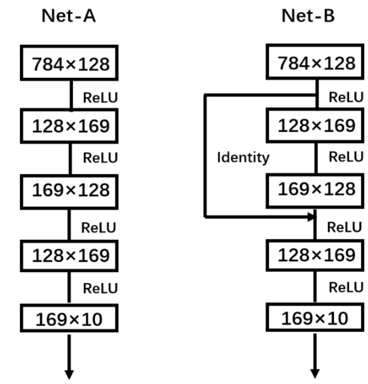


Figure 1: The parameter setting of Net-A and Net-B.

Both networks are trained on the MNIST dataset with a batch size of 100, a learning rate of 0.001, and ten epochs. We first initialize the weights of Net-A using the Xavier Initialization and set all the bias components as 0.1. As a control group, Net-B shares all the initialized parameters with Net-A. We vary the scale of the initialization before training, that is, we divide the weights from the Xavier Initialization by the 'scale'.

After training Net-A and Net-B, we calculate the $\ell_{2,2}$-norm of their weights, respectively. For Net-A, we denote the $\ell_{2,2}$-norm as $\{a, b, c, d, e\}$ in order. Similarly, we calculate $\{a', b', c', d', e'\}$ from Net-B. We obtain evidence that supports our original hypothesis by setting the scale as 10, 15, 20, 25, and other larger numbers. For each scale, we repeat the experiment fifty times.

### 2.2 Results

In Tables 1-4, we display the results for a selection of scales. $GB \triangleq abcde, GB' \triangleq a'(b'c'+1)d'e'$, where 'GB' stands for 'Generalization Bound'.
$\Delta GB \triangleq GB' - GB, \Delta acc \triangleq$ B $acc$ - A $acc$

The results suggest that with the same initialization and training strategy, the ResNet structure has a lower generalization bound than DNN (more than 95% of the data hold $a'(b'c'+1)d'e' - abcde < 0$). Since A $acc$ and B $acc$ are close, and B $acc$ is usually larger (approximately 80% of the data hold $\Delta\ acc > 0$), we conclude that the ResNets structure contributes to better generalization properties.

| $GB$ | A $acc$ | $GB'$ | B $acc$ | $\Delta GB$ | $\Delta$ $acc$(%) |
|---|---|---|---|---|---|
| 85665.89 | 0.9708 | 72200.5 | 0.9727 | -13465.4 | 0.19 |
| 113891.8 | 0.9761 | 92663.38 | 0.9779 | -21228.4 | 0.18 |
| 106829.8 | 0.9726 | 74785.51 | 0.9749 | -32044.2 | 0.23 |
| 112395.5 | 0.9718 | 70269.14 | 0.9771 | -42126.3 | 0.53 |
| 99327.02 | 0.9763 | 83384.86 | 0.9778 | -15942.2 | 0.15 |
| 102978.5 | 0.9763 | 86616.9 | 0.9735 | -16361.6 | -0.28 |
| 105634.9 | 0.977 | 85034.96 | 0.9759 | -20599.9 | -0.11 |
| 88553.07 | 0.9716 | 80143.38 | 0.9788 | -8409.69 | 0.72 |
| 115178.6 | 0.9708 | 83685.9 | 0.9764 | -31492.7 | 0.56 |
| 114659.9 | 0.9762 | 70251.56 | 0.9779 | -44408.4 | 0.17 |
| 99282.85 | 0.9712 | 73568.03 | 0.98 | -25714.8 | 0.88 |
| 146976 | 0.9743 | 81176.17 | 0.9731 | -65799.9 | -0.12 |
| 101798.5 | 0.9773 | 71842.18 | 0.9761 | -29956.3 | -0.12 |
| 117157.1 | 0.9719 | 87056.01 | 0.9733 | -30101.1 | 0.14 |
| 98932.96 | 0.9744 | 83293.12 | 0.9763 | -15639.8 | 0.19 |
| 90232.57 | 0.9738 | 80481.85 | 0.9754 | -9750.71 | 0.16 |
| 110141.9 | 0.975 | 85637.88 | 0.9747 | -24504 | -0.03 |
| 105118.1 | 0.965 | 73673.75 | 0.9752 | -31444.3 | 1.02 |
| 88778.92 | 0.9708 | 82105.02 | 0.9799 | -6673.9 | 0.91 |
| 107076.8 | 0.9737 | 81956.33 | 0.9805 | -25120.5 | 0.68 |
| 100937.7 | 0.9696 | 85669.53 | 0.9765 | -15268.2 | 0.69 |
| 108476.8 | 0.9685 | 83434.51 | 0.9759 | -25042.3 | 0.74 |
| 110783.9 | 0.9754 | 83634.4 | 0.976 | -27149.5 | 0.06 |
| 102716.6 | 0.973 | 80970.26 | 0.9768 | -21746.4 | 0.38 |
| 108043.8 | 0.971 | 83420.95 | 0.9765 | -24622.8 | 0.55 |
| 94687.19 | 0.9764 | 73752.66 | 0.9749 | -20934.5 | -0.15 |
| 102615.4 | 0.9732 | 71980.7 | 0.9744 | -30634.7 | 0.12 |
| 80745.93 | 0.974 | 72335.14 | 0.9758 | -8410.79 | 0.18 |
| 84831.58 | 0.9721 | 72048.62 | 0.9728 | -12783 | 0.07 |
| 106013 | 0.974 | 76638.36 | 0.976 | -29374.6 | 0.2 |
| 113551.4 | 0.9726 | 78923.43 | 0.9765 | -34627.9 | 0.39 |
| 97025.65 | 0.9727 | 77701.22 | 0.974 | -19324.4 | 0.13 |
| 130017.4 | 0.9759 | 70928.66 | 0.9744 | -59088.7 | -0.15 |
| 104986.7 | 0.9733 | 81410.43 | 0.9749 | -23576.3 | 0.16 |
| 113713.1 | 0.9717 | 77424.75 | 0.9772 | -36288.3 | 0.55 |
| 95741.99 | 0.9752 | 84426.64 | 0.9753 | -11315.3 | 0.01 |
| 90725.41 | 0.9713 | 73615.43 | 0.9768 | -17110 | 0.55 |
| 85339.72 | 0.9745 | 82977.7 | 0.9709 | -2362.01 | -0.36 |
| 89103.94 | 0.9732 | 77290.57 | 0.9744 | -11813.4 | 0.12 |
| 104909.9 | 0.9744 | 78088.55 | 0.9755 | -26821.3 | 0.11 |
| 142222.7 | 0.975 | 86345.11 | 0.9748 | -55877.6 | -0.02 |
| 82819.93 | 0.9745 | 73815.73 | 0.9767 | -9004.2 | 0.22 |
| 89951.42 | 0.9702 | 77795.07 | 0.9767 | -12156.4 | 0.65 |
| 112541.1 | 0.9751 | 82422.37 | 0.9766 | -30118.7 | 0.15 |
| 101843.1 | 0.9724 | 80280.58 | 0.971 | -21562.6 | -0.14 |
| 93776.51 | 0.9712 | 71715.94 | 0.9719 | -22060.6 | 0.07 |
| 102584.7 | 0.9762 | 78701.02 | 0.9765 | -23883.7 | 0.03 |
| 117927.4 | 0.9744 | 94773.64 | 0.9771 | -23153.8 | 0.27 |
| 89059.24 | 0.9732 | 76695.6 | 0.9754 | -12363.6 | 0.22 |
| 94930.14 | 0.9713 | 65332.5 | 0.9777 | -29597.6 | 0.64 |

Table 1: Scale=10

| $GB$ | A $acc$ | $GB'$ | B $acc$ | $\Delta GB$ | $\Delta$ $acc$(%) |
|---|---|---|---|---|---|
| 108746.1 | 0.9705 | 73219.09 | 0.9747 | -35527 | 0.42 |
| 168290.1 | 0.9732 | 139676.6 | 0.9766 | -28613.5 | 0.34 |
| 172926.2 | 0.9761 | 138797.1 | 0.9753 | -34129.1 | -0.08 |
| 101993.6 | 0.976 | 70107.49 | 0.9757 | -31886.1 | -0.03 |
| 134602.7 | 0.9632 | 90798.18 | 0.9733 | -43804.5 | 1.01 |
| 195945.8 | 0.9766 | 137226.7 | 0.9742 | -58719.1 | -0.24 |
| 129337.1 | 0.9755 | 117983 | 0.9759 | -11354.1 | 0.04 |
| 168796.1 | 0.9739 | 107425.9 | 0.9769 | -61370.2 | 0.3 |
| 108898.1 | 0.9723 | 67926.36 | 0.9754 | -40971.8 | 0.31 |
| 100404.4 | 0.9712 | 67844.54 | 0.9742 | -32559.9 | 0.3 |
| 100856.3 | 0.9712 | 76093.32 | 0.976 | -24763 | 0.48 |
| 131944.2 | 0.9683 | 89209.46 | 0.9784 | -42734.7 | 1.01 |
| 113163.8 | 0.9737 | 83232.78 | 0.9731 | -29931 | -0.06 |
| 223271 | 0.9757 | 125029.1 | 0.9775 | -98241.9 | 0.18 |
| 159851.6 | 0.9762 | 132680.1 | 0.9707 | -27171.4 | -0.55 |
| 161889.9 | 0.9746 | 113205.2 | 0.9785 | -48684.7 | 0.39 |
| 91169.26 | 0.9746 | 79804.93 | 0.9774 | -11364.3 | 0.28 |
| 139318.5 | 0.9745 | 114826.1 | 0.9772 | -24492.4 | 0.27 |
| 79801.2 | 0.9748 | 92302.49 | 0.9782 | 12501.3 | 0.34 |
| 147532.1 | 0.974 | 91971.79 | 0.9731 | -55560.4 | -0.09 |
| 94015.51 | 0.9719 | 90148.42 | 0.9717 | -3867.08 | -0.02 |
| 88318.33 | 0.9752 | 71322.27 | 0.9751 | -16996.1 | -0.01 |
| 129711.6 | 0.9712 | 90716.89 | 0.9748 | -38994.7 | 0.36 |
| 147862.8 | 0.975 | 99873.86 | 0.9764 | -47988.9 | 0.14 |
| 125593.9 | 0.976 | 112645.7 | 0.9743 | -12948.2 | -0.17 |
| 123896.4 | 0.9736 | 72091.85 | 0.9737 | -51804.6 | 0.01 |
| 215997.2 | 0.9735 | 141175.2 | 0.9766 | -74822.1 | 0.31 |
| 147922.2 | 0.9733 | 107571 | 0.9737 | -40351.3 | 0.04 |
| 161565.3 | 0.9757 | 77536.68 | 0.9684 | -84028.6 | -0.73 |
| 162033.8 | 0.9739 | 116951.6 | 0.9751 | -45082.2 | 0.12 |
| 109949.2 | 0.9688 | 91202.2 | 0.9768 | -18747 | 0.8 |
| 138363.3 | 0.9728 | 107571.8 | 0.9736 | -30791.5 | 0.08 |
| 166892.3 | 0.9756 | 128974.9 | 0.9797 | -37917.4 | 0.41 |
| 173282.6 | 0.9767 | 135428.4 | 0.9705 | -37854.2 | -0.62 |
| 108453.9 | 0.9764 | 96636.78 | 0.9761 | -11817.1 | -0.03 |
| 118971.6 | 0.975 | 113436.7 | 0.9737 | -5534.96 | -0.13 |
| 90118.51 | 0.9764 | 68059.32 | 0.9744 | -22059.2 | -0.2 |
| 147896.2 | 0.9727 | 121315.4 | 0.9755 | -26580.9 | 0.28 |
| 105111.5 | 0.9742 | 92290.25 | 0.973 | -12821.3 | -0.12 |
| 87627.02 | 0.9728 | 90106.48 | 0.9756 | 2479.465 | 0.28 |
| 125547.7 | 0.9768 | 107105 | 0.9789 | -18442.7 | 0.21 |
| 161633.7 | 0.9754 | 113952.3 | 0.973 | -47681.4 | -0.24 |
| 190887.7 | 0.9742 | 117740.2 | 0.9738 | -73147.5 | -0.04 |
| 146509.2 | 0.9753 | 98301.79 | 0.9747 | -48207.4 | -0.06 |
| 94129.34 | 0.9741 | 69508.67 | 0.9713 | -24620.7 | -0.28 |
| 134662.3 | 0.9736 | 119447.8 | 0.9755 | -15214.5 | 0.19 |
| 220545.9 | 0.9789 | 155478.2 | 0.9676 | -65067.7 | -1.13 |
| 121319.7 | 0.9764 | 90496.86 | 0.9758 | -30822.8 | -0.06 |
| 184569.9 | 0.9781 | 155472.6 | 0.9747 | -29097.3 | -0.34 |
| 134801.9 | 0.9747 | 91216.67 | 0.9702 | -43585.2 | -0.45 |

Table 2: Scale=15

| $GB$ | A $acc$ | $GB'$ | B $acc$ | $\Delta GB$ | $\Delta$ $acc$(%) |
|---|---|---|---|---|---|
| 206819.7 | 0.9735 | 149181.6 | 0.9788 | -57638 | 0.53 |
| 134944.9 | 0.969 | 67766.2 | 0.9742 | -67178.7 | 0.52 |
| 215699.1 | 0.9749 | 172461.6 | 0.9792 | -43237.5 | 0.43 |
| 142532 | 0.9713 | 99076.15 | 0.9731 | -43455.9 | 0.18 |
| 111196.9 | 0.9725 | 88863.99 | 0.9758 | -22332.9 | 0.33 |
| 110522.2 | 0.9736 | 85652.5 | 0.9742 | -24869.7 | 0.06 |
| 192722.1 | 0.9682 | 130982.7 | 0.9774 | -61739.4 | 0.92 |
| 162015.8 | 0.9738 | 119673.4 | 0.9745 | -42342.4 | 0.07 |
| 162374.2 | 0.9748 | 129703.1 | 0.9759 | -32671.1 | 0.11 |
| 139917 | 0.9726 | 90177.83 | 0.9764 | -49739.2 | 0.38 |
| 252037.6 | 0.9771 | 168998.1 | 0.9755 | -83039.4 | -0.16 |
| 111346.7 | 0.9732 | 104922.8 | 0.9785 | -6423.9 | 0.53 |
| 132943.7 | 0.9735 | 88195.92 | 0.9766 | -44747.8 | 0.31 |
| 121860.6 | 0.9692 | 86238.74 | 0.9738 | -35621.8 | 0.46 |
| 126338.7 | 0.9748 | 70088.84 | 0.9735 | -56249.8 | -0.13 |
| 259681.3 | 0.9764 | 137511.8 | 0.9775 | -122170 | 0.11 |
| 115183.4 | 0.9738 | 88351.95 | 0.976 | -26831.4 | 0.22 |
| 243982.3 | 0.9743 | 151450.4 | 0.9752 | -92531.8 | 0.09 |
| 175222.2 | 0.9743 | 119932.2 | 0.9794 | -55290 | 0.51 |
| 157423.5 | 0.9733 | 81937.48 | 0.975 | -75486.1 | 0.17 |
| 190885.8 | 0.9729 | 126687.3 | 0.9786 | -64198.5 | 0.57 |
| 145072.8 | 0.9767 | 83695.12 | 0.9758 | -61377.7 | -0.09 |
| 159327.9 | 0.9707 | 125827.7 | 0.9753 | -33500.2 | 0.46 |
| 180391.7 | 0.9754 | 126209.2 | 0.9741 | -54182.5 | -0.13 |
| 186597.4 | 0.9762 | 127360.8 | 0.9773 | -59236.7 | 0.11 |
| 179753.8 | 0.9739 | 89983.29 | 0.9721 | -89770.5 | -0.18 |
| 113526.2 | 0.9736 | 93217.38 | 0.9742 | -20308.8 | 0.06 |
| 264911.9 | 0.9744 | 141855.1 | 0.9772 | -123057 | 0.28 |
| 125033.6 | 0.9751 | 100900 | 0.9753 | -24133.6 | 0.02 |
| 206385.9 | 0.9781 | 151574 | 0.9763 | -54812 | -0.18 |
| 258121.2 | 0.9744 | 115381.6 | 0.9684 | -142740 | -0.6 |
| 139498.8 | 0.9757 | 132229.5 | 0.9774 | -7269.34 | 0.17 |
| 161296.5 | 0.9789 | 145447.2 | 0.972 | -15849.3 | -0.69 |
| 217909.5 | 0.976 | 121569.8 | 0.9774 | -96339.7 | 0.14 |
| 151276.1 | 0.9737 | 140211.6 | 0.9748 | -11064.6 | 0.11 |
| 230531.8 | 0.9766 | 131415.8 | 0.9778 | -99116 | 0.12 |
| 64147.79 | 0.9712 | 96333.62 | 0.9771 | 32185.83 | 0.59 |
| 116678.1 | 0.9729 | 73338.27 | 0.977 | -43339.8 | 0.41 |
| 185406.2 | 0.9765 | 119991.2 | 0.9766 | -65415 | 0.01 |
| 197436.3 | 0.9761 | 147389.9 | 0.9748 | -50046.4 | -0.13 |
| 192323.4 | 0.9736 | 147702.4 | 0.9784 | -44621 | 0.48 |
| 173957.4 | 0.9754 | 136337.6 | 0.9785 | -37619.8 | 0.31 |
| 192256.2 | 0.974 | 129720.4 | 0.9754 | -62535.9 | 0.14 |
| 118567.4 | 0.9772 | 93850.88 | 0.9785 | -24716.6 | 0.13 |
| 136173.6 | 0.973 | 112977.5 | 0.9735 | -23196.1 | 0.05 |
| 86711.27 | 0.9705 | 76979.04 | 0.9746 | -9732.22 | 0.41 |
| 146491 | 0.9725 | 133536.6 | 0.9779 | -12954.4 | 0.54 |
| 158410.3 | 0.9732 | 97535.61 | 0.9728 | -60874.7 | -0.04 |
| 96131.15 | 0.9713 | 108291.7 | 0.9753 | 12160.51 | 0.4 |
| 129003.3 | 0.9739 | 73880.03 | 0.974 | -55123.3 | 0.01 |

Table 3: Scale=20

| $GB$ | A $acc$ | $GB'$ | B $acc$ | $\Delta GB$ | $\Delta$ $acc$(%) |
|---|---|---|---|---|---|
| 129060.8 | 0.975 | 80157.64 | 0.9789 | -48903.2 | 0.39 |
| 116770 | 0.9742 | 115572.9 | 0.9756 | -1197.11 | 0.14 |
| 180311.5 | 0.9728 | 94015.31 | 0.9753 | -86296.2 | 0.25 |
| 81695.72 | 0.9712 | 97502.72 | 0.9768 | 15807 | 0.56 |
| 70339.24 | 0.969 | 96708.3 | 0.9722 | 26369.06 | 0.32 |
| 161126.5 | 0.9745 | 111803 | 0.9728 | -49323.5 | -0.17 |
| 171432 | 0.9766 | 125859.2 | 0.979 | -45572.8 | 0.24 |
| 185124.8 | 0.9733 | 111503.7 | 0.9727 | -73621.2 | -0.06 |
| 95928.25 | 0.9751 | 122182.5 | 0.974 | 26254.29 | -0.11 |
| 248100.2 | 0.975 | 150702.5 | 0.9783 | -97397.6 | 0.33 |
| 144332.9 | 0.976 | 137356.4 | 0.9756 | -6976.55 | -0.04 |
| 139907.5 | 0.9643 | 92087.47 | 0.9742 | -47820 | 0.99 |
| 197729.3 | 0.9751 | 105092.7 | 0.9702 | -92636.6 | -0.49 |
| 133659.6 | 0.9692 | 83610.34 | 0.973 | -50049.2 | 0.38 |
| 166810.5 | 0.9756 | 109556.8 | 0.9711 | -57253.8 | -0.45 |
| 215482 | 0.9753 | 106431.5 | 0.9765 | -109051 | 0.12 |
| 132191.5 | 0.9733 | 138159 | 0.9766 | 5967.517 | 0.33 |
| 229476.5 | 0.9744 | 187085.7 | 0.978 | -42390.8 | 0.36 |
| 114423.9 | 0.9702 | 68021.57 | 0.9741 | -46402.4 | 0.39 |
| 175951.7 | 0.9711 | 92180.24 | 0.9736 | -83771.4 | 0.25 |
| 227923.2 | 0.9735 | 127754.4 | 0.9727 | -100169 | -0.08 |
| 237130.8 | 0.9744 | 152768.2 | 0.9751 | -84362.6 | 0.07 |
| 179370.1 | 0.9785 | 115282 | 0.9779 | -64088.2 | -0.06 |
| 136100.1 | 0.9717 | 97372.44 | 0.971 | -38727.7 | -0.07 |
| 202884.9 | 0.972 | 147474.8 | 0.9746 | -55410.1 | 0.26 |
| 194838.9 | 0.9724 | 167291.7 | 0.9777 | -27547.2 | 0.53 |
| 226949.2 | 0.9718 | 138557.9 | 0.9734 | -88391.3 | 0.16 |
| 229911 | 0.9744 | 112083.5 | 0.975 | -117827 | 0.06 |
| 225847.8 | 0.9701 | 117439.9 | 0.971 | -108408 | 0.09 |
| 150469.4 | 0.9762 | 122022.3 | 0.9768 | -28447.1 | 0.06 |
| 199424.7 | 0.9727 | 161146.8 | 0.9751 | -38277.9 | 0.24 |
| 75988.98 | 0.9737 | 67617.81 | 0.9747 | -8371.17 | 0.1 |
| 115167 | 0.9719 | 79137.97 | 0.9737 | -36029.1 | 0.18 |
| 142186.5 | 0.9724 | 108195 | 0.975 | -33991.5 | 0.26 |
| 168483.7 | 0.9756 | 131832.8 | 0.9763 | -36650.9 | 0.07 |
| 204477.8 | 0.9735 | 131656.3 | 0.9767 | -72821.5 | 0.32 |
| 104340.1 | 0.9722 | 98541.52 | 0.9764 | -5798.63 | 0.42 |
| 166664.1 | 0.9703 | 88882.58 | 0.9739 | -77781.5 | 0.36 |
| 201134.9 | 0.9742 | 124193 | 0.9762 | -76942 | 0.2 |
| 186433.4 | 0.9718 | 121085.2 | 0.9758 | -65348.1 | 0.4 |
| 130126.4 | 0.9725 | 84957.7 | 0.9748 | -45168.7 | 0.23 |
| 125552.7 | 0.9757 | 93532.42 | 0.9749 | -32020.3 | -0.08 |
| 177650.2 | 0.9723 | 74234.16 | 0.9739 | -103416 | 0.16 |
| 151234.9 | 0.9703 | 81138.59 | 0.9776 | -70096.3 | 0.73 |
| 195276.5 | 0.9752 | 155498.7 | 0.9778 | -39777.8 | 0.26 |
| 212627 | 0.9776 | 114918.9 | 0.9753 | -97708.1 | -0.23 |
| 213248.8 | 0.9755 | 155703.2 | 0.9744 | -57545.6 | -0.11 |
| 206634.1 | 0.97 | 85726.04 | 0.9716 | -120908 | 0.16 |
| 150207.3 | 0.9733 | 66226.4 | 0.9754 | -83980.9 | 0.21 |
| 176965.5 | 0.9734 | 146542.1 | 0.9759 | -30423.4 | 0.25 |

Table 4: Scale=25

## References


[Bartlett et al.(2017)Bartlett, Foster, and Telgarsky] Bartlett, P. L., Foster, D. J., and Telgarsky, M. (2017), "Spectrally-normalized margin bounds for neural networks," in *NIPS*.

[Bousquet et al.(2003)Bousquet, Boucheron, and Lugosi] Bousquet, O., Boucheron, S., and Lugosi, G. (2003), "Concentration inequalities," *Advanced Lectures on Machine Learning: ML Summer Schools*, 208–240.

[Fan et al.(2018)Fan, Yu, and Huang] Fan, Y., Yu, J., and Huang, T. S. (2018), "Wide-activated Deep Residual Networks based Restoration for BPG-compressed Images," in *CVPR Workshops*.

[Golowich et al.(2018a)Golowich, Rakhlin, and Shamir] Golowich, N., Rakhlin, A., and Shamir, O. (2018a), "Size-Independent Sample Complexity of Neural Networks," in *Proceedings of the 31st Conference On Learning Theory*, eds. Bubeck, S., Perchet, V., and Rigollet, P., PMLR, vol. 75 of *Proceedings of Machine Learning Research*, pp. 297–299.

[Golowich et al.(2018b)Golowich, Rakhlin, and Shamir] — (2018b), "Size-Independent Sample Complexity of Neural Networks," in *COLT*.

[Goodfellow et al.(2016)Goodfellow, Bengio, and Courville] Goodfellow, I., Bengio, Y., and Courville, A. (2016), *Deep Learning*.

[He et al.(2016)He, Zhang, Ren, and Sun] He, K., Zhang, X., Ren, S., and Sun, J. (2016), "Deep Residual Learning for Image Recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778.

[Huang et al.(2017)Huang, Liu, van der Maaten, and Weinberger] Huang, G., Liu, Z., van der Maaten, L., and Weinberger, K. Q. (2017), "Densely Connected Convolutional Networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2261–2269.

[Ledoux and Talagrand(2013)] Ledoux, M. and Talagrand, M. (2013), *Probability in Banach Spaces: isoperimetry and processes*, Springer Science & Business Media.

[Li et al.(2018)Li, Lu, Wang, Haupt, and Zhao] Li, X., Lu, J., Wang, Z., Haupt, J., and Zhao, T. (2018), "On Tighter Generalization Bound for Deep Neural Networks: CNNs, ResNets, and Beyond," *arXiv preprint arXiv:1806.05159*.

[Mohri et al.(2012)Mohri, Talwalkar, and Rostamizadeh] Mohri, M., Talwalkar, A., and Rostamizadeh, A. (2012), *Foundations of machine learning (adaptive computation and machine learning series)*, Mit Press Cambridge, MA.

[Neyshabur et al.(2017)Neyshabur, Bhojanapalli, McAllester, and Srebro] Neyshabur, B., Bhojanapalli, S., McAllester, D., and Srebro, N. (2017), "A PAC-Bayesian Approach to Spectrally-Normalized Margin Bounds for Neural Networks," *CoRR*, abs/1707.09564.

[Neyshabur et al.(2015)Neyshabur, Tomioka, and Srebro] Neyshabur, B., Tomioka, R., and Srebro, N. (2015), "Norm-Based Capacity Control in Neural Networks," in *COLT*.

[Shalev-Shwartz and Ben-David(2014)] Shalev-Shwartz, S. and Ben-David, S. (2014), *Understanding Machine Learning: From Theory to Algorithms*, New York, NY, USA: Cambridge University Press.

[Srivastava et al.(2015)Srivastava, Greff, and Schmidhuber] Srivastava, R. K., Greff, K., and Schmidhuber, J. (2015), "Highway Networks," *CoRR*, abs/1505.00387.

[Sun et al.(2016)Sun, Chen, Wang, Liu, and Liu] Sun, S., Chen, W., Wang, L., Liu, X., and Liu, T.-Y. (2016), "On the Depth of Deep Neural Networks: A Theoretical View," in *AAAI*.

[Xu and Wang(2018)] Xu, Y. and Wang, X. (2018), "Understanding Weight Normalized Deep Neural Networks with Rectified Linear Units," in *Advances in Neural Information Processing Systems*, pp. 130–139.