Appendix: ProxyBNN: Learning Binarized Neural Networks via Proxy Matrices

Xiangyu He\textsuperscript{1,2}, Zitao Mo\textsuperscript{1}, Ke Cheng\textsuperscript{1,2}, Weixiang Xu\textsuperscript{1,2}, Qinghao Hu\textsuperscript{1}, Peisong Wang\textsuperscript{1}, Qingshan Liu\textsuperscript{4}, and Jian Cheng\textsuperscript{1,2,3}\textsuperscript{4}

\textsuperscript{1} NLPR, Institute of Automation, Chinese Academy of Sciences
\textsuperscript{2} School of Artificial Intelligence, University of Chinese Academy of Sciences
\textsuperscript{3} Center for Excellence in Brain Science and Intelligence Technology, Beijing, China
\{xiangyu.he, qinghao.hu, peisong.wang, jcheng\}@nlpr.ia.ac.cn
\textsuperscript{4} Nanjing University of Information Science and Technology, Nanjing, China

1 Appendix

1.1 The derivative to $Z_{i,k}$

Since the binary mapping involves all elements in a filter, the gradient of $Z_{i,k}$ requires the summation over $i = 1, \cdots, hwc$. That is

$$\frac{\partial \ell}{\partial Z_{i,k}} = \sum_{j=1}^{hwc} \frac{\partial \ell}{\partial \psi(Z_i)} \frac{\partial \psi(Z_i)}{Z_{i,k}} \sum_{n=1}^{hwc} |Z_{i,n}| \frac{\partial \text{sgn}(Z_i,j)}{\partial Z_{i,k}}$$

$$= \sum_{j=1}^{hwc} \frac{\partial \ell}{\partial \psi(Z_i)} \frac{\partial \left( \sum_{n=1}^{hwc} |Z_{i,n}| \text{sgn}(Z_{i,j}) \right)}{\partial Z_{i,k}}$$

$$= \sum_{j=1}^{hwc} \frac{\partial \ell}{\partial \psi(Z_i)} \left( \text{sgn}(Z_{i,j}) \frac{\text{sgn}(Z_{i,k})}{\text{hwc}} + \sum_{n=1}^{hwc} |Z_{i,n}| \frac{\partial \text{sgn}(Z_{i,j})}{\partial Z_{i,k}} \right)$$

$$= \frac{\text{sgn}(Z_{i,k})}{\text{hwc}} \sum_{j=1}^{hwc} \frac{\partial \ell}{\partial \psi(Z_i)} \frac{\text{sgn}(Z_{i,j})}{\text{hwc}} + \frac{\partial \ell}{\partial \psi(Z_i)} \frac{\sum_{n=1}^{hwc} |Z_{i,n}| \partial \text{sgn}(Z_{i,k})}{\text{hwc}}$$

By solving the constrained optimization problem in (3), we have

$$Z_{i,k} \approx \alpha_i B_{i,k} \approx \psi(Z_i)_k.$$  \hfill (5)

Considering an infinitesimal change in the above equation with respect to $Z_{i,k}$, we obtain

$$dZ_{i,k} \approx d \sum_{j=1}^{hwc} \frac{|Z_{i,j}|}{\text{hwc}} \text{sgn}(Z_{i,k}) + \frac{\sum_{j=1}^{hwc} |Z_{i,j}|}{\text{hwc}} d\text{sgn}(Z_{i,k})$$

which can be further simplified as

$$d\text{sgn}(Z_{i,k}) \approx \left(1 - \frac{1}{\text{hwc}} \right) \frac{hwc}{\sum_{j=1}^{hwc} |Z_{i,j}|} dZ_{i,k} \approx \frac{hwc}{\sum_{j=1}^{hwc} |Z_{i,j}|} dZ_{i,k}.$$  \hfill (7)
Combining (4) and (7), we define the derivative to $Z_{i,k}$ as

$$\frac{\partial \ell}{\partial Z_{i,k}} := \frac{\text{sgn}(Z_{i,k})}{h \cdot w \cdot c} \sum_{j=1}^{hwc} \frac{\partial \ell}{\partial \psi(Z_{i,j})} \text{sgn}(Z_{i,j}) + \frac{\partial \ell}{\partial \psi(Z_{i,k})}. \tag{8}$$