APPENDICES

A QUANTIZE RPG

Quantization refers to techniques for performing computations and storing tensors at lower bitwidths than floating point precision. Quantization can reduce model size with tiny accuracy drop. Table 9 shows that with 8-bit quantization, regular ResNet18 has accuracy drop of 0.3 percentage point, while our ResNet18-RPG has accuracy drop of 0.1 percentage point. Therefore, RPG models can be quantized for further model size reduction

Table 9: RPG model can be quantized with very tiny accuracy drop. With 8-bit quantization on ImageNet, regular ResNet18 has accuracy drop of 0.3 percentage point, while our ResNet18-RPG has accuracy drop of 0.1 percentage point.

	# Params	Acc before	Acc after ↓ quantization	Acc drop
Res18	11M	69.8	69.5	0.3
Res18-RPG		5.6M 70.2	70.1	0.1

B PROOF TO THE ORTHOGONAL PROPOSITION

We provide proofs to the orthogonal proposition mentioned in Section 3 of the main paper. Suppose we have two vectors $\mathbf{f}_i = \boldsymbol{A}_i\mathbf{f}, \mathbf{f}_j = \boldsymbol{A}_i\mathbf{f}$, where $\boldsymbol{A}_i$, $\boldsymbol{A}_j$ are sampled from the $O(M)$ Haar distribution.

Proposition 1. $\mathrm{E}\left[\langle \mathbf{f}_i, \mathbf{f}_j\rangle\right] = 0.$

Proof.

$$\begin{aligned}
\mathrm{E}\left[\langle \mathbf{f}_i, \mathbf{f}_j\rangle\right] &= \mathrm{E}\left[\langle \mathbf{f}_i, \mathbf{f}_j\rangle\right] \\
&= \mathrm{E}\left[\langle \boldsymbol{A}_i\mathbf{f}, \boldsymbol{A}_j\mathbf{f}\rangle\right] \\
&= \mathrm{E}\left[\langle \mathbf{f}, \boldsymbol{A}_i^T\boldsymbol{A}_j\mathbf{f}\rangle\right] \\
&= \mathbf{f}^T\mathrm{E}\left[\boldsymbol{A}_i^T\boldsymbol{A}_j\right]\mathbf{f} \\
&= 0
\end{aligned}$$

where $\boldsymbol{A}_i^T\boldsymbol{A}_j$ is equivalently a random sample from $O(M)$ Haar distribution and its expectation is clearly 0. □

Proposition 2. $\mathrm{E}\left[\langle \frac{\mathbf{f}_i}{\|\mathbf{f}_i\|}, \frac{\mathbf{f}_j}{\|\mathbf{f}_j\|}\rangle^2\right] = \frac{1}{M}.$

Proof.

$$\begin{aligned}
\mathrm{E}\left[\langle \frac{\mathbf{f}_i}{\|\mathbf{f}_i\|}, \frac{\mathbf{f}_j}{\|\mathbf{f}_j\|}\rangle^2\right] &= \frac{\mathrm{E}\left[\langle \boldsymbol{A}_i\mathbf{f}, \boldsymbol{A}_j\mathbf{f}\rangle^2\right]}{\|\mathbf{f}\|_2^2\|\mathbf{f}\|_2^2} \\
&= \mathrm{E}\left[\langle \boldsymbol{A}\frac{\mathbf{f}}{\|\mathbf{f}\|}, \frac{\mathbf{f}}{\|\mathbf{f}\|}\rangle^2\right], \text{where } \boldsymbol{A} = \boldsymbol{A}_i^T\boldsymbol{A}_j \sim O(M) \text{ Haar distribution}
\end{aligned}$$

Due to the symmetry,

$$= \mathrm{E}\left[\langle \boldsymbol{A}\frac{\mathbf{f}}{\|\mathbf{f}\|}, (1, 0, 0, \ldots, 0)^T\rangle^2\right]$$

Let $\mathbf{g} = \boldsymbol{A}\frac{\mathbf{f}}{\|\mathbf{f}\|}$,

$$\begin{aligned}
&= \mathrm{E}\left[g_1^2\right] \\
&= \frac{1}{M}
\end{aligned}$$

since $\mathbf{g}$ is a random unit vector and $\mathrm{E}\left[\sum_{k=1}^M g_k^2\right] = \sum_{k=1}^M \mathrm{E}\left[g_k^2\right] = 1.$ □