QuartDepth: Post-Training Quantization for Real-Time Depth Estimation on the Edge

Supplementary Material

9. Quantization with AdaRound

We adopt adaptive rounding (AdaRound) [46] for weight quantization as it performs well in post-training quantization. Specifically, different from traditional quantization with rounding-to-nearest operation, AdaRound optimizes the rounding policy so that all weights can learn the final rounding. All weights are initially rounded by floor operation. Then a learnable variable $\mathbf{v}$ is trained to determine the final rounding policy (i.e., flooring or ceiling) for each weight. The formulation can be given by

$$\hat{\mathbf{w}} = s \times \text{clip}\left(\lfloor \frac{\mathbf{w}}{s} \rfloor + \sigma(\mathbf{v}) + zp, 0, 2^k - 1\right). \tag{18}$$

$\mathbf{v}$ is the learnable parameter and the sigmoid-like function $\sigma(\cdot)$ keeps the learnable variable $\mathbf{v}$ moving between 0 and 1. The loss in quantization is formulated as following,

$$\min_{\mathbf{v}} \sum_{l=1}^{L} \left(\mathbf{w}^{(l)} - \hat{\mathbf{w}}^{(l)}\right)^{\mathrm{T}} \mathbf{F}_l \left(\mathbf{w}^{(l)} - \hat{\mathbf{w}}^{(l)}\right) + \lambda h(\mathbf{v}), \tag{19}$$

where

$$h(\mathbf{v}) = \sum_i \left(1 - |2\sigma(\mathbf{v}_i) - 1|^{\beta}\right). \tag{20}$$

We have a regularization term $h(\mathbf{v})$ in the loss to ensure that $\sigma(\mathbf{v})$ converges to either 0 or 1 with a decreasing β.

10. Additional Results

Metric3D ViT-Giant Backbone Results. We further present the results of Metric3D [95] model with ViT-Giant backbone in Table A1 on NYUv2 [47] and KITTI [19] datasets. The results show that our method achieves superior performance compared to other quantization methods especially with W4A4 configuration.

Depth Anything Full Results. We present the detailed results of Depth Anything [90] model in Table A2 with additional evaluation metrics on multiple datasets including indoor and outdoor scenes. The results show that our method achieves better performance than other two methods on nearly all evaluation metrics, especially with W4A4 configuration.

Latency Results with ViT-Giant Backbone. We provide the latency results with ViT-Giant backbone in Table A3. The results show that our quantized model achieves faster inference and higher power efficiency compared to the Float32 model. Particularly with W4A4 configuration, our method achieves 5.3× faster inference and 5.5× power efficiency.

11. Visualization Results

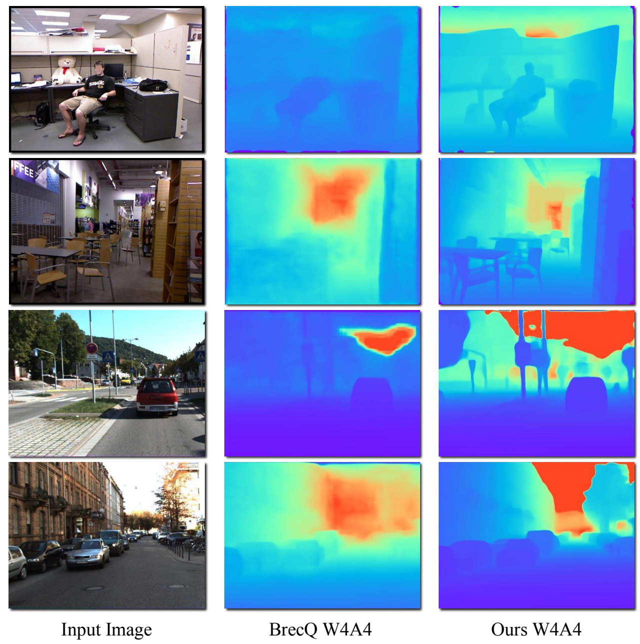

Figure A1. Visualization of the quantized Metric3D (W4A4) with ViT-Large backbone for indoor scenes (top two rows) and outdoor scenes (bottom two rows).

Method	W / A	NYUv2 [47]						KITTI [19]					
		AbsRel ↓	δ_1 ↑	δ_2 ↑	δ_3 ↑	RMSE ↓	Silog ↓	AbsRel ↓	δ_1 ↑	δ_2 ↑	δ_3 ↑	RMSE ↓	Silog ↓
ViT-Giant Backbone													
\	Float32	0.071	0.970	0.994	0.998	0.266	0.029	0.061	0.974	0.995	0.999	2.431	0.030
OBS [17]	W4	0.088	0.936	0.994	0.999	0.275	0.038	0.072	0.967	0.995	0.999	2.672	0.036
minmax [27]	W4A8	0.942	0.068	0.263	0.588	2.063	0.100	0.212	0.559	0.832	0.953	8.058	0.110
ema [26]		0.867	0.098	0.326	0.645	1.850	0.101	0.219	0.533	0.820	0.954	8.069	0.110
percentile [11]		0.926	0.079	0.282	0.597	1.986	0.098	0.231	0.501	0.786	0.940	8.325	0.112
AdaRound [46]		0.140	0.871	0.995	0.999	0.358	0.034	0.093	0.961	0.994	0.998	2.694	0.033
BrecQ [36]		0.141	0.867	0.995	0.999	0.362	0.034	0.093	0.956	0.994	0.998	2.710	0.033
Ours		**0.093**	**0.941**	**0.995**	**0.999**	**0.272**	0.035	**0.076**	**0.965**	**0.995**	**0.999**	**2.687**	**0.031**
minmax [27]	W4A4	1.844	0.062	0.145	0.287	3.808	0.161	0.407	0.277	0.481	0.632	13.911	0.237
ema [26]		1.943	0.050	0.126	0.253	3.999	0.156	0.389	0.312	0.534	0.683	13.370	0.236
percentile [11]		2.039	0.043	0.112	0.228	4.219	0.154	0.385	0.317	0.553	0.710	12.909	0.231
AdaRound [46]		0.737	0.032	0.359	0.810	1.976	0.061	0.186	0.591	0.968	0.994	4.591	0.045
BrecQ [36]		0.749	0.032	0.339	0.791	1.977	0.063	0.186	0.583	0.968	0.994	4.711	0.045
Ours		**0.119**	**0.901**	**0.994**	**0.999**	**0.333**	**0.037**	**0.068**	**0.958**	**0.993**	**0.998**	**2.938**	**0.036**

Table A1. Results of Metric3D [95] model with ViT-Giant backbone on NYUv2 and KITTI datasets.

Dataset		NYUv2 [47]								
Method	W / A	AbsRel ↓	δ_1 ↑	δ_2 ↑	δ_3 ↑	RMSE ↓	log10 ↓	RMSElog ↓	Silog ↓	SqRel ↓
\	Float32	0.056	0.984	0.998	1.000	0.206	0.024	0.072	5.277	0.017
OBS [17]	W4	0.059	0.981	0.998	1.000	0.214	0.025	0.075	5.478	0.018
BrecQ [36]	W4A8	0.099	0.903	0.994	0.999	0.439	0.046	0.128	8.563	0.059
Ours	W4A8	**0.058**	**0.982**	**0.998**	**1.000**	**0.214**	**0.025**	**0.075**	**5.461**	**0.018**
BrecQ [36]	W4A4	0.342	0.296	0.596	0.826	1.264	0.179	0.472	30.389	0.500
Ours	W4A4	**0.070**	**0.972**	**0.997**	**0.999**	**0.268**	**0.031**	**0.090**	**7.071**	**0.025**
Dataset		SUN RGB-D [67]								
\	Float32	0.500	0.660	0.960	0.980	0.616	0.088	0.259	15.483	2.175
OBS [17]	W4	0.508	0.626	0.959	0.980	0.624	0.091	0.266	15.337	2.141
BrecQ [36]	W4A8	0.489	0.692	0.961	0.981	0.597	0.084	0.252	15.498	2.149
Ours	W4A8	**0.394**	**0.756**	**0.962**	**0.983**	**0.447**	**0.076**	**0.232**	**15.086**	**1.211**
BrecQ [36]	W4A4	0.396	0.416	0.702	0.868	0.816	0.151	0.416	31.762	0.426
Ours	W4A4	**0.466**	**0.742**	**0.962**	**0.981**	**0.554**	**0.079**	**0.241**	**15.759**	2.053
Dataset		iBims-1 [32]								
\	Float32	0.150	0.714	0.966	0.991	0.593	0.073	0.185	7.515	0.130
OBS [17]	W4	0.151	0.718	0.968	0.991	0.598	0.073	0.185	7.549	0.130
BrecQ [36]	W4A8	0.183	0.593	0.941	0.980	0.764	0.092	0.227	7.679	0.195
Ours	W4A8	**0.157**	**0.700**	**0.958**	**0.986**	**0.628**	**0.077**	**0.194**	7.741	**0.144**
BrecQ [36]	W4A4	0.419	0.222	0.410	0.621	1.899	0.261	0.665	29.776	0.979
Ours	W4A4	**0.177**	**0.615**	**0.952**	**0.989**	**0.696**	**0.088**	**0.219**	**8.049**	**0.168**
Dataset		HyperSim [58]								
\	Float32	0.328	0.508	0.709	0.824	3.370	0.166	0.421	15.999	1.893
OBS [17]	W4	0.327	0.501	0.706	0.821	3.407	0.169	0.427	15.961	1.897
BrecQ [36]	W4A8	0.377	0.372	0.629	0.764	3.894	0.205	0.509	17.426	2.199
Ours	W4A8	**0.322**	**0.523**	**0.717**	**0.830**	**3.347**	**0.163**	**0.414**	**15.776**	**1.869**
BrecQ [36]	W4A4	0.703	0.047	0.103	0.179	6.291	0.549	1.317	34.756	4.684
Ours	W4A4	**0.322**	**0.512**	**0.713**	**0.828**	**3.381**	**0.166**	**0.422**	**15.999**	**1.859**
Dataset		KITTI [19]								
\	Float32	0.046	0.982	0.998	1.000	1.897	0.020	0.069	6.106	0.121
OBS [17]	W4	0.049	0.980	0.998	0.999	1.971	0.021	0.072	6.325	0.137
BrecQ [36]	W4A8	0.051	0.951	0.985	0.997	3.349	0.041	0.082	7.237	0.328
Ours	W4A8	0.060	**0.965**	**0.996**	**0.999**	**2.687**	**0.027**	0.090	8.217	**0.205**
BrecQ [36]	W4A4	0.385	0.235	0.429	0.632	12.683	0.359	1.258	8.391	3.682
Ours	W4A4	**0.059**	**0.935**	**0.981**	**0.989**	**4.194**	**0.056**	**0.093**	**7.622**	**0.425**
Dataset		vKITTI2 [7]								
\	Float32	0.084	0.912	0.986	0.995	4.008	0.039	0.138	12.096	0.430
OBS [17]	W4	0.091	0.893	0.981	0.994	4.489	0.042	0.149	12.894	0.517
BrecQ [36]	W4A8	0.606	0.132	0.275	0.447	18.052	0.734	1.721	31.342	11.434
Ours	W4A8	**0.108**	**0.846**	**0.968**	**0.992**	**5.436**	**0.052**	**0.175**	**13.870**	**0.720**
BrecQ [36]	W4A4	0.757	0.009	0.024	0.065	18.789	0.731	1.856	77.917	11.753
Ours	W4A4	**0.100**	**0.882**	**0.979**	**0.994**	**4.461**	**0.047**	**0.156**	**13.128**	**0.546**
Dataset		DIODE Outdoor [75]								
\	Float32	0.799	0.289	0.611	0.837	6.641	0.187	0.531	34.917	9.447
OBS [17]	W4	0.793	0.287	0.604	0.830	6.685	0.188	0.534	34.929	9.130
BrecQ [36]	W4A8	0.799	0.010	0.024	0.053	13.264	0.690	1.629	36.942	9.932
Ours	W4A8	0.832	**0.298**	**0.618**	**0.844**	**6.632**	**0.186**	**0.532**	**35.208**	10.520
BrecQ [36]	W4A4	0.765	0.034	0.081	0.146	13.575	0.695	1.706	58.540	10.101
Ours	W4A4	**0.758**	**0.280**	**0.605**	**0.825**	**6.801**	**0.189**	**0.533**	**35.280**	**8.470**

Table A2. Full results of Depth Anything [90] model with ViT-Large backbone.

W / A	Size (MB)	Res.	Latency (ms) ↓	FPS ↑	Power Eff. (GMAC/W)↑
ViT-Giant					
Float32	5258.5	256	1769.6 (1×)	0.57 (1×)	122.1 (1×)
W4A8	656.9		610.4 (2.9×)	1.64 (2.9×)	396.6 (3.2×)
W4A4			424.4 (4.2×)	2.36 (4.2×)	527.0 (4.3×)
Float32	5258.5	512	6446.7 (1×)	0.16 (1×)	134.1 (1×)
W4A8	656.9		2133.4 (3.0×)	0.47 (3.0×)	419.4 (3.1×)
W4A4			1425.8 (4.5×)	0.70 (4.5×)	679.1 (5.1×)
Float32	5258.5	1024	36308.1 (1×)	0.03 (1×)	142.9 (1×)
W4A8	656.9		10992.7 (3.3×)	0.09 (3.3×)	528.5 (3.7×)
W4A4			6790.3 (5.3×)	0.15 (5.3×)	790.5 (5.5×)

Table A3. Latency results of Metric3D with ViT-Giant backbone.