Mix-QViT: Mixed-Precision Vision Transformer Quantization Driven by Layer Importance and Quantization Sensitivity

Supplementary Material

Extensive Results on ImageNet1K

We evaluate Mix-QViT on ImageNet-1K [2] using ViT, DeiT, and Swin backbones under post-training quantization (PTQ) settings. As shown in Table 1, our mixed-precision scheme consistently outperforms both non-optimization and optimization-based baselines across architectures and bit-widths.

At 3-bit precision, most existing non-optimization methods degrade severely. For example, RepQ-ViT and AdaLog achieve only 0.10% and 29.42% on ViT-B, and 1.07% and 61.54% on Swin-B, respectively. LRP-AQViT improves these low-bit results substantially, reaching 52.09% on ViT-B and 70.61% on Swin-B, but Mix-QViT further increases accuracy to 56.33% and 71.27% without any optimization. With optimization, Mix-QViT achieves 79.57% on ViT-B and 80.55% on Swin-B, giving the best 3-bit accuracy across all evaluated backbones and surpassing FIMA-Q by +1.94% and +1.73% on these two models, respectively.

At 4-bit precision, Mix-QViT without optimization reaches 83.11% on ViT-B, 80.41% on DeiT-B, and 84.10% on Swin-B, improving over AdaLog's 79.68%, 78.03%, and 82.47%, respectively. Compared with LRP-AQViT, our method also provides consistent gains, improving from 81.67% to 83.11% on ViT-B, from 79.88% to 80.41% on DeiT-B, and from 83.59% to 84.10% on Swin-B. Notably, this non-optimized mixed-precision variant already exceeds several optimization-based baselines. With optimization, Mix-QViT further reaches 83.59% on ViT-B, 80.71% on DeiT-B, and 84.63% on Swin-B, outperforming FIMA-Q by +0.55%, +0.38%, and +1.03%, respectively, while delivering the best overall 4-bit results in the table.

At 6-bit precision, the baselines still show large variation, especially on smaller models. FQ-ViT, for instance, drops to 0.10 and 4.26% on ViT-S and ViT-B, respectively. More advanced calibration methods like AdaLog reach 80.91 and 84.80% on ViT-S and ViT-B, respectively. Mix-QViT yields stable, near-lossless performance across all backbones. Without optimization, it achieves 81.09% on ViT-S, 84.72% on ViT-B and 85.21% on Swin-B, matching or slightly improving over strong calibration baselines. With optimization, Mix-QViT further improves to 81.11% on ViT-S, 84.80% on ViT-B and 85.19% on Swin-B, and it outperforms all optimization-based baselines.

Extensive Results on COCO

We further evaluate Mix-QViT on COCO object detection and instance segmentation by quantizing Swin backbones in Mask R-CNN and Cascade Mask R-CNN. As shown in Table 2, Mix-QViT delivers consistent gains across 3-, 4-, and 6-bit PTQ settings and remains highly competitive even in the optimization-free regime.

At 3-bit, most optimization-free baselines degrade severely on COCO. RepQ-ViT collapses to 0.5/0.5 and 1.9/1.3 (AP^{box}/AP^{mask}) on Mask R-CNN with Swin-T and Swin-S, respectively, while AdaLog reaches only 21.0/19.4 on Mask R-CNN with Swin-S and 25.6/19.7 on Cascade Mask R-CNN with Swin-S. LRP-AQViT substantially improves these low-bit results, achieving 33.2/29.4 on Mask R-CNN with Swin-S and 37.9/31.5 on Cascade Mask R-CNN with Swin-S. Mix-QViT further improves the optimization-free setting to 31.6/29.1 and 35.4/32.3 on Mask R-CNN with Swin-T and Swin-S, and to 36.2/32.8 and 40.7/33.3 on Cascade Mask R-CNN with Swin-T and Swin-S, respectively. With optimization, Mix-QViT increases performance further to 34.5/31.1 and 38.4/35.6 on Mask R-CNN, and up to 42.2/39.8 and 45.0/41.3 on Cascade Mask R-CNN, giving the strongest 3-bit results in the table.

At 4-bit, Mix-QViT continues to outperform both fixed-bit and mixed-precision baselines. Without optimization, it achieves 44.8/40.6 and 47.6/42.7 on Mask R-CNN with Swin-T and Swin-S, and 49.6/43.0 and 51.3/44.6 on Cascade Mask R-CNN. These results surpass strong optimization-free baselines. Compared with LRP-AQViT, Mix-QViT improves from 42.9/39.9 to 44.8/40.6 on Mask R-CNN with Swin-T, from 46.8/42.2 to 47.6/42.7 on Mask R-CNN with Swin-S, and from 51.1/44.4 to 51.3/44.6 on Cascade Mask R-CNN with Swin-S. With optimization, Mix-QViT reaches 44.9/40.7 and 47.6/42.8 on Mask R-CNN, and 49.7/43.1 and 51.3/44.6 on Cascade Mask R-CNN, outperforming all optimization-based baselines by large margins.

At 6-bit, most methods approach full-precision performance, though small differences still remain. Simpler baselines such as PTQ4ViT can still collapse, while stronger methods such as AdaLog, TSPTQ-ViT, IGQ-ViT, and MPTQ-ViT recover results close to full precision. In the optimization-free setting, Mix-QViT achieves 45.9/41.5 and 48.3/43.2 on Mask R-CNN with Swin-T and Swin-S, and 50.2/43.6 and 51.8/44.9 on Cascade Mask R-CNN. With optimization, Mix-QViT further reaches 50.4/43.8 on

Table 1. Quantization results for image classification on ImageNet-1K [2]. Each value reports Top-1 accuracy (%). "Optim." denotes optimization-based methods; "Prec. (W/A)" indicates bit precision for weights and activations. "MP" is mixed precision; "$*$" indicates that the model size and bit operations are equivalent to those of a fixed-bit quantized model; "†" = replicated results. Bold indicates best performance. Note that the reference numbers in the supplementary materials are not the same as in the main paper.

Method	Opti.	Prec.(W/A)	ViT-S	ViT-B	DeiT-T	DeiT-S	DeiT-B	Swin-S	Swin-B
Full-Precision	-	32/32	81.39	84.54	72.21	79.85	81.80	83.23	85.27
RepQ-ViT [5]†	×	3/3	0.10	0.10	0.10	3.27	7.57	1.37	1.07
AdaLog [18]†	×	3/3	12.63	29.42	25.70	22.82	55.90	58.12	61.54
LRP-AQViT [13]	×	MP3/MP3	18.88	52.09	31.33	52.43	67.22	68.12	70.61
Mix-QViT (ours)*	×	MP3/MP3	**24.19**	**56.33**	**38.12**	**57.51**	**70.09**	**71.80**	**71.27**
BRECQ [3]	✓	3/3	0.42	0.59	25.52	14.63	46.29	11.67	1.7
QDrop [16]	✓	3/3	4.44	8.00	30.73	22.67	24.37	60.89	54.76
PD-Quant [9]	✓	3/3	1.77	13.09	39.97	29.33	0.94	69.67	64.32
I&S-ViT [22]	✓	3/3	45.16	63.77	41.52	55.78	73.30	74.20	69.30
DopQ-ViT [20]	✓	3/3	54.72	65.76	44.71	59.26	74.91	74.77	69.63
FIMA-Q [19]	✓	3/3	64.09	77.63	55.55	69.13	76.54	77.26	78.82
Mix-QViT (ours)*	✓	MP3/MP3	**69.13**	**79.57**	**58.97**	**72.08**	**77.92**	**79.98**	**80.55**
PTQ4ViT [21]	×	4/4	42.57	30.69	36.96	34.08	64.39	76.09	74.02
APQ-ViT [1]	×	4/4	47.95	41.41	47.94	43.55	67.48	77.15	76.48
RepQ-ViT [5]	×	4/4	65.05	68.48	57.43	69.03	75.61	79.45	78.32
AdaLog [18]	×	4/4	72.75	79.68	63.52	72.06	78.03	80.77	82.47
LRP-AQViT [13]	×	MP4/MP4	74.72	81.67	64.70	75.83	79.88	81.76	83.59
Mix-QViT (ours)*	×	MP4/MP4	**75.82**	**83.11**	**65.49**	**76.56**	**80.41**	**82.20**	**84.10**
ERQ [23]	✓	4/4	71.61	78.65	61.79	74.35	79.18	81.19	83.32
I&S-ViT [22]	✓	4/4	74.87	80.07	65.21	75.81	79.97	81.17	82.60
DopQ-ViT [20]	✓	4/4	75.69	80.95	65.54	75.84	80.13	81.71	83.34
FIMA-Q [19]	✓	4/4	76.68	83.04	66.84	76.87	80.33	81.82	83.60
Mix-QViT (ours)*	✓	MP4/MP4	**79.18**	**83.59**	**68.37**	**77.50**	**80.71**	**82.41**	**84.63**
FQ-ViT [8]	×	6/6	4.26	0.10	58.66	45.51	64.63	66.50	52.09
PSAQ-ViT [4]	×	6/6	37.19	41.52	57.58	63.61	67.95	72.86	76.44
Ranking-ViT [11]	×	6/6	-	75.26	-	74.58	77.02	-	-
PTQ4ViT [21]	×	6/6	78.63	81.65	69.68	76.28	80.25	82.38	84.01
APQ-ViT [1]	×	6/6	79.10	82.21	70.49	77.76	80.42	82.67	84.18
RepQ-ViT [5]	×	6/6	80.43	83.62	70.76	78.90	81.27	82.79	84.57
AdaLog [18]	×	6/6	80.91	84.80	71.38	79.39	81.55	83.19	85.09
IGQ-ViT [12]	×	6/6	80.76	83.77	71.15	79.28	81.71	82.86	84.82
Mix-QViT (ours)*	×	MP6/MP6	**81.09**	**84.72**	**71.78**	**79.59**	**81.60**	**83.27**	**85.21**
EasyQuant [17]	✓	6/6	75.13	81.42	-	75.27	79.47	82.45	84.30
NoisyQuant-Linear [10]	✓	6/6	76.86	81.90	-	76.37	79.77	82.78	84.57
BRECQ [3]	✓	6/6	54.51	68.33	70.28	78.46	80.85	82.02	83.94
QDrop [16]	✓	6/6	70.25	75.76	70.64	77.95	80.87	82.60	84.33
PD-Quant [9]	✓	6/6	70.84	75.82	70.49	78.40	80.52	82.51	84.32
Bit-shrinking [6]	✓	6/6	80.44	83.16	-	78.51	80.47	82.44	-
I&S-ViT [22]	✓	6/6	80.43	83.82	70.85	79.15	81.68	82.89	84.94
DopQ-ViT [20]	✓	6/6	80.52	84.02	71.17	79.30	81.69	82.95	84.97
FIMA-Q [19]	✓	6/6	80.64	84.82	71.53	79.52	81.74	83.19	85.01
Mix-QViT (ours)*	✓	MP6/MP6	**81.11**	**84.80**	**71.85**	**79.63**	**81.74**	**83.34**	**85.19**

Cascade Mask R-CNN with Swin-T and 51.9/44.9 with Swin-S. Overall, these results show that Mix-QViT maintains stable, near-lossless performance at 6-bit precision while remaining highly competitive with the strongest existing PTQ methods.

References

[1] Yifu Ding, Haotong Qin, Qinghua Yan, Zhenhua Chai, Junjie Liu, Xiaolin Wei, and Xianglong Liu. Towards accurate post-training quantization for vision transformer. In *Proc. 30th ACM Int. Conf. Multimedia*, pages 5380–5388, 2022. 2, 3

Table 2. Quantization results for object detection and instance segmentation on COCO [7]. AP^{box} and AP^{mask} denote box and mask average precision, respectively. "Prec. (W/A)" specifies bit precision for weights and activations. "MP" denotes mixed precision; "$*$" indicates fixed-bit equivalent cost (model size and bit operations); "†" indicates replicated results. Bold indicates best performance. Note that the reference numbers in the supplementary materials are not the same as in the main paper.

Method	Optim.	Prec. (W/A)	Mask R-CNN				Cascade Mask R-CNN			
			w.Swin-T		w.Swin-S		w.Swin-T		w.Swin-S	
			AP^{box}	AP^{mask}	AP^{box}	AP^{mask}	AP^{box}	AP^{mask}	AP^{box}	AP^{mask}
Full-Precision		32/32	46.0	41.6	48.5	43.3	50.4	43.7	51.9	45.0
RepQ-ViT [5]†	×	3/3	0.5	0.5	1.9	1.3	0.7	0.7	1.3	1.2
AdaLog [18]†	×	3/3	12.6	11.4	21.0	19.4	20.8	15.6	25.6	19.7
LRP-AQViT [13]	×	MP3/MP3	28.2	26.1	33.2	29.4	33.2	**31.1**	37.9	31.5
Mix-QViT (ours)*	×	MP3/MP3	**31.6**	**29.1**	**35.4**	**32.3**	**36.2**	**32.8**	**40.7**	**33.3**
Mix-QViT (ours)*	✓	MP3/MP3	**34.5**	**31.1**	**38.4**	**35.6**	**42.2**	**39.8**	**45.0**	**41.3**
RepQ-ViT [5]	×	4/4	36.1	36.0	44.2	40.2	47.0	41.4	49.3	43.1
AdaLog [18]	×	4/4	39.1	37.7	44.3	41.2	48.2	42.3	50.6	44.0
TSPTQ-ViT [14]	×	4/4	42.9	39.3	45.0	40.7	47.8	41.6	48.8	42.5
IGQ-ViT [12]	×	4/4	41.0	38.8	44.8	41.3	48.5	42.4	50.5	44.0
MPTQ-ViT [15]	×	MP4/MP4	44.2	40.2	47.3	**42.7**	49.2	42.7	50.8	44.2
LRP-AQViT [13]	×	MP4/MP4	42.9	39.9	46.8	42.2	49.3	**43.0**	51.1	44.4
Mix-QViT (ours)*	×	MP4/MP4	**44.8**	**40.6**	**47.6**	**42.7**	**49.6**	**43.0**	**51.3**	**44.6**
I&S-ViT [22]	✓	4/4	37.5	36.6	43.4	40.3	48.2	42.0	50.3	43.6
DopQ-ViT [20]	✓	4/4	37.5	36.5	43.5	40.4	48.2	42.1	50.3	43.7
FIMA-Q [19]	✓	4/4	38.7	37.8	44.2	41.1	48.7	42.5	50.4	43.7
Mix-QViT (ours)*	✓	MP4/MP4	**44.9**	**40.7**	**47.6**	**42.8**	**49.7**	**43.1**	**51.3**	**44.6**
PTQ4ViT [21]	×	6/6	5.8	6.8	6.5	6.6	14.7	13.6	12.5	10.8
APQ-ViT [1]	×	6/6	45.4	41.2	47.9	42.9	48.6	42.5	50.5	43.9
RepQ-ViT [5]	×	6/6	45.1	41.2	47.8	43.0	50.0	43.5	51.4	44.6
AdaLog [18]	×	6/6	45.4	41.3	48.0	**43.2**	50.1	43.6	51.7	44.8
TSPTQ-ViT [14]	×	6/6	45.8	41.4	**48.3**	**43.2**	50.2	43.5	51.8	44.8
IGQ-ViT [12]	×	6/6	45.5	**41.5**	48.2	**43.2**	**50.4**	**43.8**	**51.9**	45.0
MPTQ-ViT [15]	×	MP6/MP6	**45.9**	41.4	**48.3**	43.1	50.2	43.6	51.8	44.8
Mix-QViT (ours)*	×	MP6/MP6	**45.9**	**41.5**	**48.3**	**43.2**	50.2	43.6	51.8	**44.9**
Mix-QViT (ours)*	✓	MP6/MP6	**45.9**	**41.5**	**48.3**	**43.2**	**50.4**	**43.8**	**51.9**	**44.9**

[2] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet classification with deep convolutional neural networks. *Adv. Neural Inform. Process. Syst. (NeurIPS)*, 25, 2012. 1, 2

[3] Yuhang Li, Ruihao Gong, Xu Tan, Yang Yang, Peng Hu, Qi Zhang, Fengwei Yu, Wei Wang, and Shi Gu. Brecq: Pushing the limit of post-training quantization by block reconstruction. *arXiv:2102.05426*, 2021. 2

[4] Zhikai Li, Liping Ma, Mengjuan Chen, Junrui Xiao, and Qingyi Gu. Patch similarity aware data-free quantization for vision transformers. In *Eur. Conf. Comput. Vis. (ECCV)*, pages 154–170, 2022. 2

[5] Zhikai Li, Junrui Xiao, Lianwei Yang, and Qingyi Gu. RepQ-ViT: Scale reparameterization for post-training quantization of vision transformers. In *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, pages 17227–17236, 2023. 2, 3

[6] Chen Lin, Bo Peng, Zheyang Li, Wenming Tan, Ye Ren, Jun Xiao, and Shiliang Pu. Bit-shrinking: Limiting instantaneous sharpness for improving post-training quantization. In *Proc. IEEE/CVF Conf. Comput. Vis. and Pattern Recog. (CVPR)*, pages 16196–16205, 2023. 2

[7] Tsung-Yi Lin et al. Microsoft COCO: Common objects in context. In *Eur. Conf. Comput. Vis. (ECCV), Zurich, Switzerland: Springer*, pages 740–755, 2014. 3

[8] Yang Lin, Tianyu Zhang, Peiqin Sun, Zheng Li, and Shuchang Zhou. FQ-ViT: Post-training quantization for fully quantized vision transformer. In *Proc. 31st Int. Joint Conf. Artif. Intell. (IJCAI)*, pages 1173–1179, 2022. 2

[9] Jiawei Liu, Lin Niu, Zhihang Yuan, Dawei Yang, Xinggang Wang, and Wenyu Liu. Pd-quant: Post-training quantization based on prediction difference metric. In *Proc. IEEE/CVF Conf. Comput. Vis. and Pattern Recog. (CVPR)*,

pages 24427–24437, 2023. 2

[10] Yijiang Liu, Huanrui Yang, Zhen Dong, Kurt Keutzer, Li Du, and Shanghang Zhang. Noisyquant: Noisy bias-enhanced post-training activation quantization for vision transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20321–20330, 2023. 2

[11] Zhenhua Liu, Yunhe Wang, Kai Han, Wei Zhang, Siwei Ma, and Wen Gao. Post-training quantization for vision transformer. *Adv. Neural Inform. Process. Syst. (NeurIPS)*, 34: 28092–28103, 2021. 2

[12] Jaehyeon Moon, Dohyung Kim, Junyong Cheon, and Bumsub Ham. Instance-aware group quantization for vision transformers. In *Proc. IEEE/CVF Conf. Comput. Vis. and Pattern Recog. (CVPR)*, pages 16132–16141, 2024. 2, 3

[13] Navin Ranjan and Andreas Savakis. LRP-QViT: Mixed-precision vision transformer quantization using layer importance score. In *2025 25th International Conference on Digital Signal Processing (DSP)*, pages 1–5. IEEE, 2025. 2, 3

[14] Yu-Shan Tai, Ming-Guang Lin, and An-Yeu Andy Wu. TSPTQ-ViT: Two-scaled post-training quantization for vision transformer. In *IEEE Int. Conf. Acoust., Speech and Sig. Process. (ICASSP)*, pages 1–5, 2023. 3

[15] Yu-Shan Tai et al. MPTQ-ViT: Mixed-precision post-training quantization for vision transformer. *arXiv:2401.14895*, 2024. 3

[16] Xiuying Wei, Ruihao Gong, Yuhang Li, Xianglong Liu, and Fengwei Yu. Qdrop: Randomly dropping quantization for extremely low-bit post-training quantization. *arXiv preprint arXiv:2203.05740*, 2022. 2

[17] Di Wu, Qi Tang, Yongle Zhao, Ming Zhang, Ying Fu, and Debing Zhang. Easyquant: Post-training quantization via scale optimization. *arXiv preprint arXiv:2006.16669*, 2020. 2

[18] Zhuguanyu Wu, Jiaxin Chen, Hanwen Zhong, Di Huang, and Yunhong Wang. AdaLog: Post-training quantization for vision transformers with adaptive logarithm quantizer. In *Eur. Conf. Comput. Vis. (ECCV)*, pages 411–427, 2025. 2, 3

[19] Zhuguanyu Wu, Shihe Wang, Jiayi Zhang, Jiaxin Chen, and Yunhong Wang. FIMA-Q: Post-training quantization for vision transformers by fisher information matrix approximation. In *Proc. IEEE/CVF Conf. Comput. Vis. and Pattern Recog. (CVPR)*, pages 14891–14900, 2025. 2, 3

[20] Lianwei Yang, Haisong Gong, Haokun Lin, Yichen Wu, Zhenan Sun, and Qingyi Gu. Dopq-vit: Towards distribution-friendly and outlier-aware post-training quantization for vision transformers. *arXiv:2408.03291*, 2024. 2, 3

[21] Zhihang Yuan, Chenhao Xue, Yiqi Chen, Qiang Wu, and Guangyu Sun. PTQ4ViT: Post-training quantization for vision transformers with twin uniform quantization. In *Eur. Conf. Comput. Vis. (ECCV)*, pages 191–207, 2022. 2, 3

[22] Yunshan Zhong, Jiawei Hu, Mengzhao Chen, Rongrong Ji, et al. I&s-vit: An inclusive & stable method for pushing the limit of post-training vits quantization. *arXiv:2311.10126*, 2023. 2, 3

[23] Yunshan Zhong, You Huang, Jiawei Hu, Yuxin Zhang, and Rongrong Ji. Towards accurate post-training quantization of vision transformers via error reduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025. 2